LE SCIENZE
edizione italiana di SCIENTIFIC AMERICAN

numero 24
agosto 1970
anno III
volume V

Armi chimiche e biologiche

Gli Stati Uniti hanno rinunciato a ogni tipo di armi biologiche e a far uso per primi della maggior parte delle armi chimiche. Resta aperta la questione se includere o meno nella proibizione gas irritanti ed erbicidi

di Matthew S. Meselson

Il 25 novembre dello scorso anno il presidente Nixon ha annunciato una serie di importanti decisioni riguardanti le armi chimiche e biologiche. Egli ha dichiarato che gli Stati Uniti non faranno mai uso per primi di armi chimiche letali o inabilitanti e che non useranno mai in nessuna circostanza armi biologiche, neanche come ritorsione. Egli ha anche annunciato che avrebbe sottoposto al Senato il Protocollo di Ginevra del 1925, che proibisce l'uso di armi chimiche e biologiche in guerra.

La decisione del presidente lasciava in sospeso l'interrogativo se le tossine, le sostanze velenose ma inorganiche prodotte da alcuni batteri e altri organismi, fossero comprese nella rinuncia totale alle armi biologiche o solo nell'impegno a non usare per primi quelle chimiche. La domanda trovò risposta il 14 febbraio, quando il presidente estese il criterio delle armi biologiche fino a comprendere anche le tossine. Allo scopo di eliminare ogni possibile ambiguità, fu deciso che la rinuncia all'uso delle armi a base di tossine venisse preventivamente estesa anche al caso in cui i progressi della chimica consentissero di prepararle sinteticamente anziché per estrazione da batteri o altri microorganismi.

In seguito alla nuova politica, gli USA distruggeranno le scorte esistenti di armi a base di germi e di tossine e non si impegneranno piú nel loro sviluppo, produzione o accumulazione. Il programma americano verrà limitato alla ricerca per scopi rigorosamente difensivi, per esempio per le tecniche di immunizzazione.

Per quanto riguarda le armi chimiche, invece, le restrizioni sono meno illimitate. Gli Stati Uniti si riservano il diritto di proseguire lo sviluppo, la produzione e l'accumulazione di armi chimiche letali o inabilitanti. Si impegnano tuttavia a non usarle mai per primi. Lo stesso impegno è stato incluso nel progetto di Trattato da parte delle 84 nazioni che ora fanno parte del Protocollo di Ginevra.

Ci sono due tipi di armi chimiche al cui uso bellico non si è rinunciato; essi sono le sostanze chimiche erbicide e gli agenti per il « controllo dei tumulti », come il CS, impiegati dalle forze armate americane nel sud-est asiatico. Gli Stati Uniti hanno asserito che tali agenti non sono proibiti dal Protocollo di Ginevra. Molte nazioni non sono d'accordo su questa questione, e ciò introduce dei problemi per la ratifica di quel trattato da parte americana. Lo stesso presidente non ha fatto alcun riferimento al CS o agli erbicidi, mentre i funzionari della Casa Bianca hanno detto solo che l'uso di queste armi « per il momento » continuerà.

La decisione del presidente era basata su una indagine di sei mesi sulla politica americana e sui programmi per le armi chimiche e biologiche. L'indagine era stata svolta sotto gli auspici del Consiglio per la sicurezza nazionale, che coordinava i lavori di vari uffici governativi.

L'argomento fu sottoposto a un esame minuzioso. Il Consiglio per la sicurezza nazionale prese in considerazione i seguenti temi: le armi chimiche e biologiche che erano già pronte e quelle che potevano essere presumibilmente prodotte, la possibile importanza militare di ciascun tipo di armi in relazione ad altre armi disponibili, le conseguenze eventuali della proliferazione, i problemi strategici relativi all'uso di tali armi, l'effetto politico e morale del loro sviluppo e dell'eventuale uso, i rischi potenziali per le popolazioni civili, le difese possibili contro i vari tipi di armi, la politica e l'attività svolta dalle altre nazioni.

Era ovvio, e lo divenne sempre piú nel corso dell'indagine, che tale inchiesta era fortemente in ritardo. Gli Stati Uniti possedevano un grande impianto per la produzione di armi batteriologiche, anche se nessuno aveva ancora definito uno « scenario » convincente per il caso in cui esse dovessero essere adoperate. In breve, sia prima che durante l'indagine, le armi chimiche furono implicate in una preoccupante serie di incidenti, creando un'atmosfera favorevole a chiarimenti sulla politica attuata nel passato. Una serie di terremoti che scossero la città di Denver fu fatta risalire alla presenza di grossi depositi sotterranei di rifiuti chimici presso il Rocky Mountain Arsenal dell'esercito.

Aeroplani dell'aviazione americana spruzzano erbicidi sulle terre del Vietnam. Dal 1962 sono state compiute piú di 19 000 missioni di questo tipo. Un aeroplano è in grado di coprire una striscia di terreno larga circa 100 metri e lunga 18 chilometri. I militari affermano che la distruzione dei raccolti riduce le scorte di viveri delle forze nemiche mentre la distruzione del fogliame nella giungla riduce le possibilità di occultamento.

Lo scarico accidentale di gas nervini durante la prova di un serbatoio spruzzatore d'aereo sopra il Poligono di Dugway, nell'Utah, provocò la morte di centinaia di pecore, alcune delle quali pascolavano a quasi 70 chilometri di distanza. In seguito vi fu la protesta contro l'esercito che progettava di trasportare 12 000 tonnellate di bombe a gas nervino di tipo antiquato, attraverso tutti gli USA dal Colorado fino all'Atlantico, dove dovevano essere affondate. Subito dopo ci fu l'incidente di Okinawa, dove il gas nervino sfuggí da un deposito americano di munizioni, e in quella occasione fu anche rivelata l'esistenza di depositi di gas. Infine ci fu una serie di articoli dedicati al fatto che il Dipartimento della difesa stava sperimentando nel Pacifico meridionale delle armi biologiche e che aveva svolto prove di gas nervino in aria aperta nelle Hawai senza che gli organi ufficiali ne fossero informati.

La reazione pubblica a questi incidenti può essere valutata dal fatto che al Senato americano una risoluzione, che richiedeva il riesame da parte del Servizio della sanità pubblica di qualunque programma che prevedesse il trasporto interno o le prove all'aperto di agenti chimici letali o di qualunque agente per la guerra biologica, fu approvata con 91 voti favorevoli e nessun contrario. La risoluzione, divenuta ora legge, richiede inoltre che il Dipartimento della difesa fornisca al Congresso rapporti semestrali sulle ricerche, lo sviluppo, le prove e l'immagazzinamento di tutti gli agenti per la guerra chimica o biologica.

In questo articolo analizzeremo in primo luogo la natura delle armi chimiche e biologiche (CBW) e le loro implicazioni militari, poi i tentativi internazionali che sono stati compiuti per prevenirne l'uso, e infine la situazione attuale. Per quanto gran parte del lavoro riguardante le CBW sia stato segreto, gli elementi essenziali che riguardano le armi sono ora liberamente disponibili. In particolare va segnalato il rapporto steso da un gruppo di lavoro dell'ONU e pubblicato l'anno scorso, e inoltre il successivo rapporto dell'Organizzazione mondiale della sanità (WHO), che esamina specificamente gli effetti potenziali sulle popolazioni civili.

Ancora piú dettagliato dei rapporti dell'ONU e della WHO è infine lo studio svolto dall'Istituto di Ricerche per la Pace Internazionale di Stoccolma, che sarà pronto fra breve.

Per quanto le armi chimiche e biologiche siano strettamente collegate nella psicologia, nelle consuetudini e nelle leggi internazionali che ne limitano l'uso, se si devono analizzare caratteristiche e implicazioni militari è utile distinguere tali armi in diverse categorie. Esamineremo cinque tipi di armi: armi biologiche letali, armi biologiche inabilitanti, armi chimiche letali, armi chimiche inabilitanti e agenti antipiante. La distinzione fra letale e inabilitante non è affatto ben definita, particolarmente nelle condizioni estremamente prive di controllo di una guerra. Esiste una vasta gamma di agenti che vanno da quelli fortemente mortali a quelli in genere non mortali, cui si aggiungono gas lacrimogeni da usare assieme alle pallottole e alle bombe per accrescere il numero di vittime. Tuttavia gli argomenti a favore e contro le armi letali e non letali meritano una specifica attenzione.

Prendiamo in considerazione innanzi tutto le armi biologiche letali. Esse dovrebbero funzionare mediante disseminazione di nubi di germi patogeni sopra la zona da colpire o sopravento rispetto a essa. I germi verrebbero cosí inspirati dagli esseri viventi che occupano la zona bersaglio. L'antrace è un esempio di malattia che può essere provocata in questo modo. Essa è causata da un batterio chiamato *Bacillus anthracis*, e colpisce soprattutto gli animali domestici, trasmettendosi qualche volta all'uomo. Dato che essa non è molto contagiosa fra gli esseri umani, i casi naturali di antrace vengono in genere localizzati e non producono epidemie. Se però i batteri venissero spruzzati in aria sotto forma di aerosol, gli effetti potrebbero essere disastrosi. Si valuta che l'inalazione di circa 50 000 spore del *B. anthracis* (il cui peso totale è inferiore a un milionesimo di grammo) sia sufficiente a determinare 50 probabilità su cento di contrarre l'antrace polmonare. I primi sintomi si manifesterebbero circa un giorno dopo l'attacco. All'inizio la malattia potrebbe essere scambiata per un comune raffreddore, ma in seguito sopravverrebbero forte tosse, cianosi, incapacità di respirazione e morte. L'antrace polmonare, se non curato, è quasi sempre mortale.

Il rapporto WHO fornisce stima sull'entità delle aree che possono essere coperte da diversi agenti biologici. Come esempio illustrativo il rapporto suppone che un bombardiere leggero scarichi in un solo passaggio delle « bombolette » che funzionano al suolo, contenenti in tutto 50 chilogrammi di agenti essiccati e in polvere, per una lunghezza di due chilometri e ad angolo retto rispetto al vento. Si è supposta una intensità di turbolenza atmosferica inferiore a un certo livello, ma non tanto bassa da essere del tutto eccezionale, specialmente di notte. Le bombolette liberano l'agente sotto forma di aerosol, che si solleva quindi su tutta l'area colpita. La valutazione tiene conto dell'efficacia di disseminazione delle bombolette, della velocità di decadimento dell'infettività dell'agente, del grado di diluizione verticale nell'atmosfera, della velocità di deposizione al suolo e della curva dose-effetto per l'uomo. Per un attacco a base di spore di antrace il rapporto WHO prevede un elevato grado di mortalità su un'area di almeno 20 chilometri quadrati. Per quanto vi siano dei dubbi sui dati che intervengono nei calcoli, la valutazione è volutamente prudente. Il rapporto dell'ONU contiene una valutazione analoga: esso prende in considerazione un attacco a bassa quota realizzato da parte di un grande bombardiere, il quale distribuisca 10 000 chilogrammi di agente per una lunghezza di 100 chilometri mediante un serbatoio a spray. In base a questa valutazione l'area nella quale si verificherebbe un'alta percentuale di vittime giunge a una estensione di 100 000 chilometri quadrati, a seconda del particolare agente adoperato.

Fra gli agenti biologici mortali che potrebbero essere presi in considerazione per uso militare vanno annoverati i virus dell'encefalite equina orientale e della febbre gialla, le rickettsie, che provocano la febbre delle Montagne Rocciose, e i batteri dell'antrace, della peste, del colera e della morva. Esistono vaccini e antibiotici limitatamente efficaci contro alcune di queste malattie, ma non contro tutte. Inoltre tali protezioni possono essere annullate da un attacco massiccio, mentre l'azione di alcuni agenti antibiotici può venire elusa facendo uso di varietà di agenti patogeni che siano resistenti agli antibiotici. Ci si può proteggere mediante maschere antigas o rifugi provvisti di filtro d'aria, sempre però che si disponga di un sistema di preallarme degli attacchi, mentre nessun sistema pienamente soddisfacente è stato ancora realizzato. In ogni caso un piano per fornire alla popolazione civile maschere e rifugi e per mantenere la disciplina necessaria per il loro uso richiederebbe un grosso e prolungato sforzo sia economico che politico, senza che a ciò corrisponda una protezione contro l'attacco biologico realmente valida.

Da queste brevi considerazioni dovrebbe risultare chiaro che le armi biologiche letali potrebbero rappresentare una minaccia paurosa di distruzione per la popolazione umana sopra aree molto vaste. La minaccia viene resa particolarmente grave dalla relativa facilità con cui molti possono venire in possesso di tali armi non appena si siano impadroniti della relativa tecnologia e dalla possibilità che cambino le consuetudini e le tendenze che finora hanno tenuto le nazioni lontane dalla tentazione di sviluppare le armi biologiche. Dopo diversi anni spesi nel perseguire un programma di sviluppo delle armi biologiche, programma che non ha avuto da parte dell'esecutivo adeguato controllo o guida, e che è stato quasi completamente nascosto alla conoscenza sia dell'opinione pubblica sia dei rappresentanti al Congresso, gli Stati Uniti hanno deciso che il modo migliore per ridurre al minimo la minaccia delle armi biologiche è di rinunziare del tutto a esse. La giustificazione logica di tale decisione sta nella convinzione che il possesso di tali armi accrescerebbe di poco o nulla la capacità americana di deterrente strategico, mentre la loro proliferazione rappresenterebbe una grossa minaccia per gli Stati Uniti e per tutta l'umanità.

Il secondo tipo di armi da prendere in considerazione è costituito dalle armi biologiche inabilitanti. L'encefalite equina venezuelana, per esempio, è prodotta da un virus che provoca forti emicranie, nausee e prostrazione, ma che presenta in epidemie naturali una percentuale inferiore allo 0,5 per cento di casi di morte. I metodi per disseminare gli agenti biologici inabilitanti e le possibilità di difesa contro di essi sono essenzialmente gli stessi degli agenti biologici letali. L'uso di armi biologiche inabilitanti dovrebbe essere preso in considerazione solo in certe situazioni poco frequenti ed esterne. Lo scenario forse piú « attraente » che sia stato proposto immagina che si sia verificato un accerchiamento di un grosso contingente di forze alleate da parte di truppe nemiche spiegate su di una area molto estesa e mescolate alla popolazione civile. In tale situazione l'impiego di armi biologiche inabilitanti nelle zone anteriori e posteriori impedisce l'avanzata del nemico abbastanza a lungo da consentire il rinforzo o l'evacuazione delle truppe alleate, nel caso in cui le uniche alternative dovessero consistere o nell'uso di armi nucleari o nella resa.

Evidentemente ogni decisione di lanciare un attacco biologico a base di armi inabilitanti deve affrontare il fatto che un notevole numero di civili, particolarmente bambini e malati, resterebbe quasi inevitabilmente ucciso, anche se la percentuale di morte acciden-

tale fosse solo di pochi decimi percento. Oltre a ciò c'è la possibilità che nelle condizioni esistenti in zona di guerra la probabilità di morte accidentale possa risultare molto piú alta di quella stimata durante i casi di manifestazione spontanea della malattia e da vari tipi di dati sperimentali. È anche importante notare che tale scenario prende in considerazione un caso di *primo* uso di armi biologiche piuttosto che il loro impiego come ritorsione contro un analogo attacco del nemico. Ciò aggiunge il rischio addizionale dell'escalation e dell'allargamento del conflitto che potrebbero derivare dall'esplosione di una guerra biologica con germi di ogni tipo. A lunga scadenza, un grave attentato alla sicurezza dovuto all'uso di armi biologiche inabilitanti, o anche al solo fatto di conservarle, sarebbe rappresentato dallo stimolo alla proliferazione delle armi batteriologiche, comprese quelle letali. Le attrezzature necessarie per realizzare, produrre e portare a destinazione le armi biologiche inabilitanti sono essenzialmente le stesse richieste per i germi letali. Le leggi e le consuetudini internazionali non fanno distinzioni fra di esse.

Per concludere, le armi biologiche inabilitanti potrebbero sembrare logiche in certe situazioni tattiche, ma tali situazioni sono molto poco frequenti, e inoltre i rischi potrebbero superare i possibili vantaggi.

A questo punto bisognerebbe fare una precisazione piú generale sulle armi batteriologiche. Tali armi presentano delle serie carenze da un punto di vista militare. I loro effetti non sono prevedibili come quelli delle altre armi, e potrebbero sfuggire al controllo. Armi alternative e già esistenti sono preferibili, e la realizzazione di una forza strategica basata sulle armi biologiche finirebbe per essere una aggiunta e non un'alternativa ai preesistenti programmi di spese militari.

Le autorità militari e i leader politici manifestano una forte avversione all'uso delle armi biologiche, in parte per ragioni pratiche, in parte per la scarsa familiarità con esse e in parte per il rifiuto morale e il timore sempre associati a queste armi. Attualmente non sembra che esista il minimo serio interesse per le armi biologiche nelle alte sfere militari. Per quanto tali armi possano divenire una terribile minaccia, la probabilità che ciò si verifichi è stata considerevolmente ridotta dallo spettacolare gesto di rinuncia compiuto dagli USA. Questo è un momento storico, nel quale le scienze biologiche, intimamente legate alla missione umanitaria della medicina, possono venire sottratte in tutto il mondo all'asservimento agli scopi militari. Se poi il disarmo biologico internazionale potrà essere o meno assicurato per un lungo periodo, può in ultima analisi dipendere dai progressi nel disarmo chimico, dato che a giudizio di molti i due problemi sono connessi.

Le armi chimiche rappresentano un problema molto piú immediato. Esse furono adoperate durante la I guerra mondiale e attualmente sono immagazzinate da parte degli USA e dell'URSS, ma le posseggono anche molte altre nazioni. La distinzione fra le armi letali e quelle inabilitanti ha una importanza piú pratica in questo caso che non in quello delle armi biologiche. Esistono notevoli resistenze all'uso delle armi chimiche letali, e fino a non molto tempo fa anche quelle non letali erano state escluse dall'uso bellico. Dal punto di vista del controllo degli armamenti sarebbe molto utile trattare assieme sia le armi chimiche letali che quelle non letali. Tuttavia, allo scopo di rendere piú chiara possibile l'analisi, è conveniente considerarle separatamente.

Le moderne armi chimiche letali adoperano i gas nervini prodotti per la prima volta dalla Germania durante la II guerra mondiale (ma non adoperati). Questi agenti sono centinaia di volte piú potenti dei gas tossici della I guerra mondiale: uccidono non appena inalati o non appena si depositano sulla pelle sotto forma di goccioline. Il termine « gas nervino » deriva dal fatto che questi agenti operano interferendo con la trasmissione degli impulsi nervosi attraverso le sinapsi. Essi inattivano l'enzima colinesterasi, che normalmente funziona per porre fine alla trasmissione di un impulso nervoso. In presenza di un agente nervino gli impulsi nervosi proseguono senza alcun controllo, provocando il collasso della respirazione e di altre funzioni. La morte provocata da intossicazione di gas nervino si verifica per asfissia; è preceduta da annebbiamento della vista, salivazione e convulsioni.

Le riserve americane comprendono due tipi di agenti nervini. Uno è il GB, sigla di codice che sta a indicare l'isopropilmetilfosfonoflorurato; noto anche come Sarin, è stato prodotto in piccole quantità dalla Germania durante la II guerra mondiale. Il Sarin è un liquido volatile che evapora a temperatura ambiente sotto forma di gas incolore e inodore.

Le armi che contengono il Sarin lo liberano sotto forma di spray, che poi evapora rappresentando un pericolo per la respirazione per chiunque non sia opportunamente protetto. Si valuta che il grado di esposizione letale per gli esseri umani sia approssimativamente di 100 milligrammi al metro cubo al minuto. Ciò vuol dire, per esempio, che un uomo esposto all'aria con una concentrazione di Sarin di 10 milligrammi al metro cubo assorbirebbe una dose mortale in dieci minuti.

Dato che il pericolo offerto dal Sarin è soprattutto respiratorio, basta una maschera antigas a conferire una buona protezione. Le maschere antigas moderne sono capaci di ridurre fino a circa centomila volte la concentrazione di tutti i gas bellici noti. In piú per gli agenti nervini esistono antidoti chimici che sono efficaci quando la dose dell'agente non è molto forte e l'antidoto viene somministrato in tempo.

L'altro tipo di agente nervino disponibile nei depositi americani è il VX. La formula chimica del VX è ancora segreta, per quanto il rapporto WHO suggerisca che si tratti dell'etil-S-dimetilamminoetilmetilfosfonotiolato. Questo fa parte di una serie di composti preparati per la prima volta verso la metà degli anni 50 nel corso di una ricerca per ottenere migliori insetticidi (anche il Sarin scaturí da una ricerca sugli insetticidi). Il VX è anch'esso liquido ma è molte volte piú tossico del Sarin e molto meno volatile; è mortale sia se viene respirato sia se si deposita sulla pelle. Il VX uccide nello spazio di pochi minuti; inoltre, contaminando il terreno e gli oggetti sui quali si è depositato, è in grado di rendere pericolosa un'intera zona per molti giorni. È stato il VX che ha ucciso le pecore nell'Utah.

La dose letale di VX applicato alla pelle è stata valutata fra i 2 e i 10 milligrammi, a seconda della zona colpita. Dato che anche il solo contatto con una gocciolina di VX può essere mortale, per una adeguata protezione è necessario indossare una apposita tuta protettiva oltre che una maschera antigas, il che provoca notevole impaccio: tale abbigliamento rende infatti difficoltosi i movimenti mentre lo smaltimento del calore diviene un serio problema. L'efficienza bellica risulterebbe fortemente ridotta nel caso in cui si debba indossare un equipaggiamento protettivo completo, cosí come dalla stretta osservanza di varie speciali precauzioni necessarie per sopravvivere in ambiente chimico letale. Tale caratteristica di complessità della guerra di gas fu chiaramente riconosciuta già durante la I guerra mondiale. Un ufficiale cosí si esprimeva: « Esistono infiniti problemi. Come può il soldato mangiare, bere, dormire, svolgere le funzioni corporee, usare le proprie armi, dare e ricevere ordini? Come può sapere quando la zona a cui è vicino è contaminata? ».

Per uso tattico contro un nemico privo di equipaggiamenti protettivi le armi chimiche sarebbero distruttive. Contro un nemico che disponga di tute e di maschere e in grado di far indossare tale equipaggiamento alle proprie truppe e in piú con la minaccia di rendere pan per focaccia, le armi chimiche letali complicherebbero enormemente le battaglie senza fornire a nessuna delle due parti alcun vantaggio. Questo ragionamento suggerisce di non dare inizio a guerre chimiche letali. Esso, tuttavia, suggerisce anche una ragione per possedere le armi chimiche letali come deterrente tattico nel caso si ritenga che l'avversario le possegga a sua volta. Se per esempio in Europa scoppiasse una guerra terrestre convenzionale, e se uno solo degli avversari disponesse di munizioni chimiche letali, questi potrebbe essere tentato di usarle allo scopo di costringere le truppe avversarie a dei ripiegamenti difensivi mentre le proprie forze, avvantaggiate dalla conoscenza del tempo e del luogo degli attacchi chimici, passerebbero all'offensiva. Una volta in possesso di questi elementi, sempre secondo il suddetto ragionamento, le loro operazioni sarebbero molto meno complicate. È una analisi di questo tipo che sta presumibilmente alla base della politica americana di conservare le armi chimiche letali nonostante il principio enunciato sia quello di non adoperarle mai per primi.

In realtà, si possono mettere in dubbio sia la plausibilità che l'esattezza del precedente scenario. Molti sostengono che una grande guerra è estremamente improbabile in Europa. Ed è anche piú improbabile che una tale guerra possa poi svolgersi per molti giorni senza che si faccia ricorso alle armi nucleari, nel qual caso quelle chimiche perderebbero ogni importanza. E infine, si dice, anche nel caso di una grande guerra ancora non affrontata con armi nucleari, l'uso di quelle chimiche sarebbe fortemente sconsigliabile per il rischio che, usando tali armi non convenzionali allo scopo di ottenere qualche notevole vantaggio, si faccia scattare una risposta nucleare.

Esistono ovviamente alcuni rischi in entrambi i modi di vedere il problema. Il secondo modo, tuttavia, consentirebbe la rinuncia al possesso delle armi chimiche letali.

Le nazioni non nucleari potrebbero produrre armi chimiche letali allo scopo di procurarsi capacità d'attacco strategico su popolazioni urbane. In condizioni meteorologiche favorevoli all'attacco, un bombardiere medio o un aereo da trasporto commerciale opportunamente adattato può liberare una quantità di agente nervino sufficiente a uccidere una forte percentuale di popolazione non protetta nella zona centrale di una grande città. Per esempio, il rapporto WHO valuta che un attacco condotto dall'aria, controvento, per uno sviluppo di due chilometri e in cui vengano scaricate quattro tonnellate di agente chimico sopra una città, provocherebbe forte mortalità entro un'area compresa fra 2 e 40 chilometri quadrati, a seconda del tipo di agente e di armi adoperato. Disponendo di un opportuno preallarme una popolazione estremamente disciplinata potrebbe difendersi da tale attacco facendo uso combinato di maschere antigas, di rifugi protettivi e di una terapia a base di antidoti. Mentre le armi chimiche strategiche non aggiungono niente di

AGENTI	MALATTIE	PERIODO DI INCUBAZIONE (GIORNI)	EFFETTO DEL TRATTAMENTO	CONTAGIOSITÀ
VIRUS	ENCEFALITE EQUINA ORIENTALE	DA 5 A 15	NESSUNO	PER VETTORE
	ENCEFALITE DA ACARO	DA 7 A 14	NESSUNO	PER VETTORE
	FEBBRE GIALLA	DA 3 A 6	NESSUNO	PER VETTORE
RICKETTSIE	FEBBRE DELLE MONTAGNE ROCCIOSE	DA 3 A 10	BUONO	PER VETTORE
	TIFO EPIDEMICO	DA 6 A 15	BUONO	PER VETTORE
BATTERI	ANTRACE	DA 1 A 5	MODESTO	BASSA
	COLERA	DA 1 A 5	BUONO	ALTA
	PESTE POLMONARE	DA 2 A 5	MODESTO	ALTA
	TULAREMIA	DA 1 A 10	BUONO	BASSA
	TIFO	DA 7 A 21	BUONO	ALTA

AGENTI	MALATTIE	PERIODO DI INCUBAZIONE (GIORNI)	EFFETTO DEL TRATTAMENTO	CONTAGIOSITÀ
VIRUS	FEBBRE CHIKUNGUNYA	DA 2 A 6	NESSUNO	PER VETTORE
	FEBBRE DENGUE	DA 5 A 8	NESSUNO	PER VETTORE
	ENCEFALITE EQUINA VENEZUELANA	DA 2 A 5	NESSUNO	PER VETTORE
RICKETTSIE	FEBBRE Q	DA 10 A 21	BUONO	BASSA
BATTERI	BRUCELLOSI	DA 7 A 21	MODESTO	NESSUNA
FUNGHI	COCCIDIO-IDOMICOSI	DA 7 A 21	SCARSO	NESSUNA

In questa tabella vengono elencati diversi agenti biologici; in alto sono riportati quelli che possono provocare la morte, in basso quelli che provocano inabilità. Il contagio « per vettore » indica la trasmissione per mezzo di certi tipi di zanzare o di altri insetti.

NOME COMUNE	SARIN	VX	IPRITE	BZ	CS E CS-2
CLASSIFICAZIONE MILITARE	AGENTE LETALE (GAS NERVINO)	AGENTE LETALE (GAS NERVINO)	AGENTE LETALE E INABILITANTE	AGENTE INABILITANTE (PSICOCHIMICO)	AGENTE AGGRESSIVO
STATO FISICO	LIQUIDO	LIQUIDO	LIQUIDO	SOLIDO	SOLIDO
FORMA IN CUI VIENE DISSEMINATO	VAPORE, AEROSOL O SPRAY	AEROSOL O SPRAY	SPRAY	AEROSOL O POLVERE	AEROSOL O POLVERE
DURATA DELLA CONTAMINAZIONE	ORE O GIORNI	GIORNI O SETTIMANE	GIORNI O SETTIMANE		CS: MINUTI CS-2: SETTIMANE

Agenti per guerra chimica contenuti nei magazzini americani elencati per categoria e secondo alcune delle loro caratteristiche. Questi dati sono stati raccolti dall'Organizzazione Mondiale della Sanità nel suo rapporto sulle armi chimiche e biologiche.

significativo agli arsenali delle potenze nucleari, la proliferazione di tali armi fra le nazioni non nucleari costituirebbe invece un serio pericolo. Anche se è infatti poco probabile che una piccola nazione sia in grado di sferrare un attacco chimico sopra una vasta zona in un paese che disponga di moderne difese aeree, sarebbe però molto piú facile penetrare nello spazio aereo di una o piú città costiere.

È anche importante considerare l'eventuale ruolo delle armi chimiche letali nei conflitti locali. Oggi tali conflitti vengono combattuti con armi a esplosivo di grande potenza e a fiamma, che individualmente hanno un'area di efficacia limitata. Sebbene tali guerre possano essere esageratamente distruttive, esse divengono tali solo quando vengono adoperate enormi quantità di armi (per esempio nel 1968 gli USA in Vietnam adoperarono piú di 6000 tonnellate di munizioni al giorno).

Molti tipi di munizioni adoperate in una guerra locale potrebbero però essere riempiti con mezzi chimici letali. In tal caso l'« area letale » di munizioni leggere quali proiettili da mortaio e razzi potrebbe venire accresciuta anche di un fattore 100. Anche se i combattenti fossero muniti di equipaggiamento protettivo, tali armi sarebbero distruttive sia per le unità militari colte impreparate sia per gli abitanti delle aree urbane. Piccole unità militari acquisterebbero cosí capacità strategiche contro città.

È chiaro perciò che le armi chimiche costituiscono una minaccia ben al di là di quella derivante dal loro possibile impiego da parte delle maggiori potenze. È tale minaccia che rappresenta uno degli argomenti principali a favore del disarmo chimico e biologico. Un freno definitivo, ma su cui spesso si passa sopra, circa l'uso tattico delle armi chimiche letali, particolarmente quando il campo di battaglia si trova su suolo alleato, consiste nel fatto che il loro uso in larga scala potrebbe provocare una forte mortalità fra i civili indifesi sia nella zona di combattimento sia anche a notevole distanza sottovento rispetto a tale zona.

Per quanto la politica americana consideri attualmente le tossine alla stregua delle armi biologiche, rinunciando anche al loro possesso, il rapporto dell'ONU classifica le tossine come sostanze chimiche, in quanto non si riproducono. Le tossine sono sostanze velenose prodotte da organismi viventi che comprendono piante, animali e batteri. Alcuni esempi sono la ricina, estratta dal seme di ricino, la tetrodotossina estratta da un pesce della famiglia dei Diodontidi (pesce istrice) e la botulina prodotta dal batterio *Clostridium botulinum*. Alcune tossine, come la botulina, sono estremamente mortali per l'uomo; altri, come l'enterotossina (stafilococco responsabile degli avvelenamenti da cibo guasto), producono solo temporanee indisposizioni.

Anche se le tossine non sono in grado di riprodursi e quindi non possono causare epidemie, esse producono molti degli stessi sintomi associati con infezioni prodotte da organismi patogeni. Infatti i principali sintomi patologici di diverse malattie dovute a batteri sono probabilmente provocati da tossine prodotte nel corpo umano dai microorganismi viventi. Le armi a base di tossine sono quindi strettamente connesse a quelle biologiche.

Ai fini dell'uso bellico, le tossine potrebbero venire disperse sotto forma di aerosol quasi nello stesso modo delle armi biologiche e chimiche. Dato che le tossine in pratica non vengono assorbite attraverso la pelle, le maschere antigas potrebbero fornire una sufficiente protezione, e cosí anche i rifugi muniti di speciali filtri d'aria. Una protezione si potrebbe ottenere anche mediante una precedente immunizzazione con tossoidi specifici. Ogni tossoide, però, è efficace solo contro un particolare tipo di tossina, e per alcune tossine il margine di protezione non è sufficiente per essere di qualche utilità pratica.

La principale ragione militare per disporre di tossine consiste nel fatto che, a causa della loro grande potenza, il peso delle munizioni a base di tossine necessario per coprire una determinata area sarebbe minore di quello corrispondente richiesto con normali munizioni chimiche. Ci sono diverse ragioni tecniche per nutrire dubbi su ciò, ma se anche fosse cosí il risparmio che ne risulterebbe sarebbe molto poco importante per grandi potenze militari con grosse capacità logistiche. Un efficace programma americano per armi a base di tossine sarebbe d'altra parte andato contro la decisione del presidente di demilitarizzare e rendere di pubblico dominio la ricerca e gli impianti di produzione per le armi biologiche americane, e sarebbe stato impossibile per gli USA assumere una posizione non equivoca e convincente contro l'uso di malattie come arma da guerra.

Le prime armi chimiche a essere impiegate durante la I guerra mondiale erano di tipo non letale. Risulta che alcuni soldati portarono al fronte dei candelotti a gas lacrimogeno in dotazione della polizia. Ben presto sia la Germania che la Francia cominciarono ad adoperare proiettili d'artiglieria contenenti gas lacrimogeno, e migliaia di

questi proiettili furono sparati diversi mesi prima del famoso attacco tedesco di Ypres durante il quale venne liberato un gas a base di cloro contenuto entro cilindri. Gas lacrimogeni e altre sostanze chimiche irritanti seguitarono a essere usati durante tutta la guerra, per un totale di piú di 12 000 tonnellate. Durante la II guerra mondiale erano state preparate dai belligeranti di entrambe le parti quantità anche maggiori di tali sostanze chimiche, che non furono però adoperate.

Le moderne armi chimiche inabilitanti sono di due tipi, uno i cui effetti si manifestano molto piú a lungo del periodo di esposizione e uno di effetto limitato. La rinuncia da parte del presidente Nixon dell'uso per primo delle sostanze chimiche inabilitanti è stata estesa solo al tipo a effetto prolungato. Un esempio di questo tipo è l'agente BZ, americano. Si tratta di una sostanza psicochimica, la cui composizione chimica è ancora segreta, anche se il rapporto WHO suppone che esso appartenga alla famiglia dei cosiddetti benzilati. Il BZ è una sostanza solida che può essere dispersa sotto forma di aerosol in modo da essere inspirata dai nemici; agisce sia sui processi fisici che su quelli mentali, provocando annebbiamento della vista, disorientamento e confusione. I suoi effetti inabilitanti possono perdurare per diversi giorni.

Il BZ, anche se è stato messo in produzione come arma dall'esercito USA ed è stato inserito in diverse munizioni, non viene considerato agente inabilitante molto soddisfacente. Esso può provocare un comportamento imprevedibile e spesso violento. Si potrebbe verificare che sotto la sua influenza, uomini che già abbiano sufficienti ragioni per combattere possano farlo in maniera molto piú accanita. Inoltre il BZ ha seri effetti sul bilancio idrico del corpo e sul meccanismo di regolazione della temperatura, effetti che potrebbero portare alla morte, particolarmente in condizioni di ambiente caldo e secco. Molti sforzi sono stati dedicati, peraltro senza successo, alla ricerca di un inabilitante a effetto prolungato che non presenti questi inconvenienti, o altri analoghi.

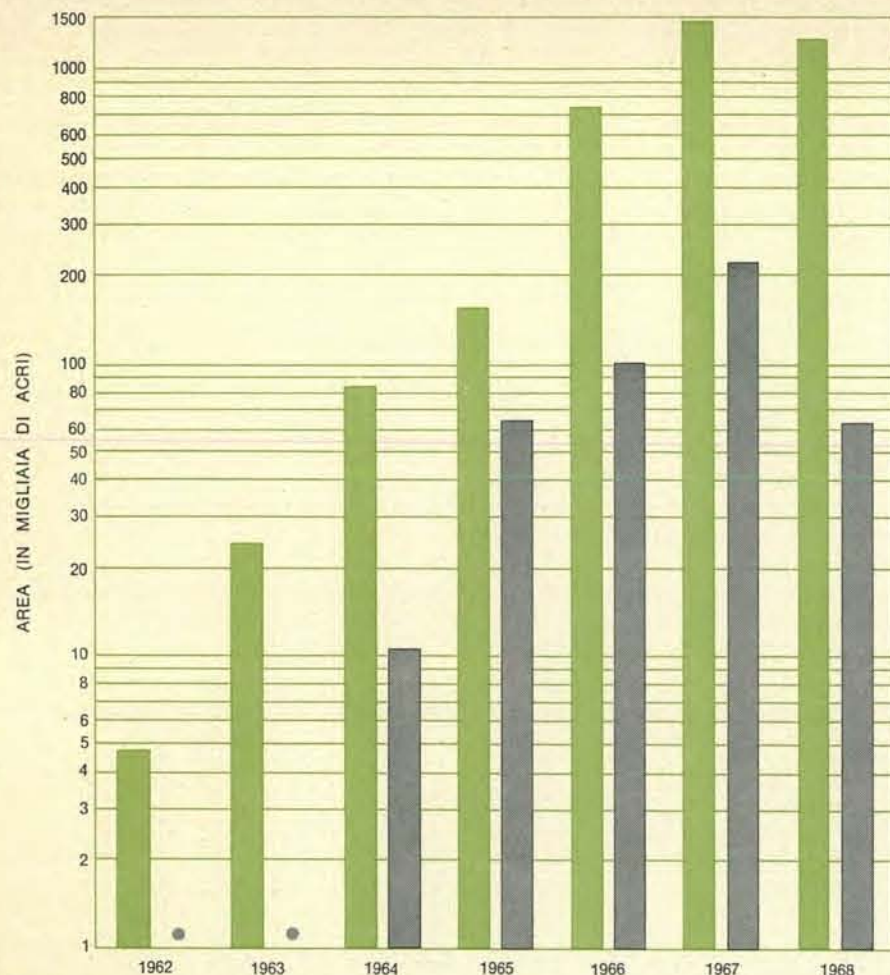
Il principale inabilitante a effetto limitato attualmente in uso militare è il CS (ortoclorobenzalmalonitrile). Tale composto fu prodotto per la prima volta negli USA negli anni 20, con processo di sintesi. Dopo la II guerra mondiale è stato sviluppato dal governo inglese come agente per il controllo di tumulti e denominato dalle iniziali dei suoi scopritori americani, Ben Corson e Roger Staughton del Middlebury College. In relazione al suo impiego per scopi militari esso viene piú precisamente definito come un agente aggressivo.

Le prime munizioni a base di CS funzionavano vaporizzando l'agente mediante miscela pirotecnica. Il CS poi condensa in modo da formare un aerosol costituito da particelle delle dimensioni del micron. Il CS pirotecnico viene adoperato in granate, razzi, proiettili d'artiglieria e bombe a grappolo. Un tipo piú nuovo è denominato CS-2. Usato sia in meccanismi di disseminazione sia in bombe dirompenti, esso consiste di una polvere formata da particelle delle dimensioni del micron trattata con gel di silice e siliconi per migliorare le sue caratteristiche di movimento e di persistenza nell'ambiente. Il CS-2 può restare attivo nella zona d'azione per diverse settimane, e viene rimesso in circolo dal vento e dal movimento della gente e dei veicoli.

L'effetto del CS dipende dalle dimensioni delle particelle dell'aerosol. Particelle di dimensioni superiori ai 50 micron esercitano un effetto predominante sugli occhi, mentre particelle di dimensioni minori sono piú efficaci come irritanti dei polmoni. Il CS-2 per uso militare viene ridotto a dimensioni tali da provocare questo ultimo effetto. L'esposizione a entrambe le forme di CS provoca forte dolore agli occhi e al tratto superiore delle vie respiratorie, propagandosi poi fino agli intimi recessi dei polmoni e dando luogo a sensazioni di soffocamento e di acuta ansietà. L'esposizione moderatamente intensa della pelle in aria umida può provocare la formazione di piaghe che richiedono molti giorni per guarire.

Se l'esposizione non è eccessiva, i sintomi passano in genere entro pochi minuti dal termine dell'esposizione. La dose letale per l'uomo, da come è stata stimata in base a esperimenti sugli animali, è di gran lunga superiore a quella richiesta per provocare intensa irritazione. Tuttavia una esposizione intensa o prolungata, come si potrebbe avere in zone adiacenti o nelle strette vicinanze di una postazione dalla quale si sprigioni l'aggressivo, potrebbe causare seri danni ai polmoni e anche la morte, particolarmente fra bambini e malati. Nessun effetto ritardato invece hanno finora rilevato le esposizioni moderate al CS; per quanto siano attualmente in corso indagini su tale eventualità, queste non sono state pe-

	TIPO DI ARMA		
AGENTI	NUCLEARE	CHIMICA	BIOLOGICA
AREA COLPITA	FINO A 300 CHILOMETRI QUADRATI	FINO A 60 CHILOMETRI QUADRATI	FINO A 100 000 CHILOMETRI QUADRATI
TEMPO RICHIESTO PER L'EFFETTO	SECONDI	MINUTI	GIORNI
DANNI ALLE STRUTTURE	VASTA DISTRUZIONE	NESSUNO	NESSUNO
ALTRI EFFETTI	PROLUNGATA RADIOATTIVITÀ ENTRO UN'AREA DI 2500 CHILOMETRI QUADRATI	CONTAMINAZIONE PER GIORNI O SETTIMANE	EVENTUALE EPIDEMIA O NUOVI FOCOLAI DI MALATTIA
RITORNO ALL'USO NORMALE	DA 3 A 6 MESI	USO LIMITATO PER GIORNI O SETTIMANE	VARIABILE
EFFETTO SULL'UOMO	90 % DI MORTI	50 % DI MORTI	50 % DI MALATI

Danni provocati da tre diversi tipi di arma: una bomba atomica da un megaton, 15 tonnellate di agente nervino e 10 tonnellate di agente biologico. In tutti e tre i casi si è assunto che l'arma venga sganciata da un bombardiere su una popolazione indifesa.

Dati sull'uso di erbicidi nel Vietnam del Sud, forniti dall'aviazione americana, relativi alla distruzione del fogliame (*in colore*) e dei raccolti (*in grigio*). I punti rappresentano meno di 1000 acri (un acro è pari a circa mezzo ettaro). Nel Vietnam del Sud circa 4 milioni di ettari sono coltivati mentre 7 milioni sono ricoperti da foreste.

rò ancora completate.

Dal punto di vista militare il CS ha soppiantato i vecchi agenti aggressivi come l'ordinario gas lacrimogeno (il CN, ossia il cloroacetofenone) e l'agente emetico e respiratorio noto come adamsite o DM.

Quando vengono adoperati in guerra contro gente sprovvista di maschera, gli agenti aggressivi sono efficaci nel costringere il nemico a uscire dal riparo per affrontare la cattura o il fuoco avversario, nel restringergli il terreno o nel disturbare il suo fuoco. Contro personale dotato di maschera gli agenti aggressivi sono molto meno efficaci, sebbene riducano alquanto l'efficienza di combattimento obbligando gli uomini a mettersi la maschera. La prima importante applicazione in combattimento di gas aggressivo dopo la II guerra mondiale è avvenuta nel Vietnam, dove le forze armate americane hanno fatto uso di più di 7 milioni di chilogrammi di queste sostanze. Agli inizi la politica degli Stati Uniti era quella di far uso del CS solo quando il suo impiego fosse risultato più « umano » dell'uso di armi più micidiali. Per esempio, il 24 marzo 1965, facendo seguito al primo servizio giornalistico sull'uso di gas non letali in Vietnam da parte degli USA, il segretario di stato Dean Rusk fece la seguente dichiarazione: « Non riteniamo che i gas verranno usati in ordinarie operazioni militari... La previsione è naturalmente, che queste armi verranno adoperate solo in quelle situazioni che implichino controllo di tumulti o situazioni analoghe a quelle di controllo di tumulti ».

Per cinque mesi dopo la dichiarazione di Rusk l'uso di agenti aggressivi nel Vietnam cessò del tutto. Poi si verificò un fatto che mise l'uso del CS nella luce più attraente possibile e probabilmente giocò un ruolo importante nel far rilasciare l'autorizzazione per il suo rinnovato impiego. Il 5 settembre un ufficiale della marina raggiunse una caverna dove si riteneva che con i civili si trovassero mescolati soldati vietcong. Dovendo scegliere se mandare le forze all'assalto, se lanciare dentro delle granate di CS o se abbandonare la missione, egli decise per il CS. Si è saputo poi che vennero fuori diversi soldati nemici assieme a 400 civili, senza che questi ultimi subissero alcun danno. Dopo questo incidente soldati e comandanti di campo trovarono una grande varietà di usi per il CS in combattimento. I soldati si trovarono in molte situazioni in cui esso poteva essere usato per produrre vittime tra il nemico e d'altra parte per compiere la missione riducendo le proprie perdite. Uno dei più importanti usi del CS in Vietnam consiste nel provocare l'uscita dei soldati nemici dai bunker prima di iniziare un fuoco massiccio o un assalto di fanteria.

L'utilità complessiva del CS in Vietnam non è nota, in quanto non sono stati fatti studi sistematici in proposito. Nondimeno è stata un'arma molto diffusa, e in seguito a tale pressione proveniente dall'area di guerra il suo uso è stato esteso notevolmente. Un indice di questo fatto è reperibile nella documentazione degli acquisti dell'esercito per il sud-est asiatico: le forniture sono aumentate da 114 000 chilogrammi nell'anno fiscale 1965 a 2 720 000 chilogrammi nel 1969. Un altro indice è la rapida proliferazione di munizioni, sia sperimentali che recentemente standardizzate dall'esercito, a base di CS. Appena nel luglio 1966 i manuali militari elencavano solo cinque tipi di armi a base di CS: due granate e tre congegni a disseminazione di contenuto. Da allora sono apparse 18 nuove munizioni CS, che vanno dalle granate munite di detonatore a tempo e destinate a essere lanciate con paracadute, fino a proiettili per obice da 105 e da 155 millimetri, a varie munizioni per mortaio e per razzo, e a diverse bombe d'aereo capaci di coprire aree fino a un chilometro quadrato.

È stato altre volte sostenuto che le armi chimiche non letali potrebbero rendere più « umana » la guerra. C'è invece una buona ragione per aspettarsi l'opposto. Finché in guerra vengono adoperate armi mortali, l'introduzione di quelle chimiche non letali ovviamente non vorrà dire che queste verranno impiegate da sole ma piuttosto in appoggio alle armi già in servizio, allo scopo di accrescere l'efficacia complessiva delle operazioni militari. Senza dubbio ciò si è verificato nel caso dell'agente CS.

L'estensione dell'uso del CS nel Vietnam ha fatto sorgere due ampie questioni. La prima è che gli USA hanno abbandonato la primitiva giustificazione umanitaria per l'uso in guerra di sostanze chimiche. La seconda è che diventa sempre più chiaro che il costo del controllo a lunga scadenza delle armi può essere per gli USA molto gravoso. Su ciò ritorneremo in seguito.

L'ultima categoria di armi chimiche che prenderemo in considerazione è quella degli agenti antipianta. Questi agenti vennero sviluppati per la prima volta con scopi militari durante la II guerra mondiale e in seguito furono largamente usati per la distruzione delle erbacce. Verso la fine della guerra venne presa in considerazione la possibilità di adoperarli per distruggere il riso che veniva coltivato dai soldati giapponesi asserragliati in isole fuori mano, ma l'esecuzione del piano non ricevette mai l'autorizzazione. In Malesia gli inglesi fecero uso in piccola scala di erbicidi per pulire il fondo stradale durante le loro operazioni militari in quel paese negli anni 50.

L'uso di erbicidi nel Vietnam fu autorizzato la prima volta in via sperimentale nel 1961. In seguito essi cominciarono a essere adoperati per aumentare la visibilità lungo strade e vie d'acqua e lungo i bordi perimetrali delle installazioni militari; poi per la distruzione di raccolti che si pensava fossero destinati al consumo da parte del nemico, e infine per il trattamento di vaste aree alberate che si sospettava nascondessero campi base nemici o strade di rifornimento. Verso la metà del 1969 erano stati cosparsi circa 20 000 chilometri quadrati, di cui il 10% di terra coltivata. Dopo un massimo nel 1967, le operazioni contro la terra coltivata erano state notevolmente ridotte, il che rispecchiava la discutibilità sia della loro efficacia che della loro opportunità.

Sono stati adoperati tre principali agenti o miscele defolianti designati come Arancione, Bianco e Blu. L'agente Arancione è una miscela in parti eguali degli n-butil-esteri dell'acido 2,4-diclorofenossiacetico (2,4-D) e dell'acido 2,4,5-triclorofenossiacetico (2,4,5-T). Tale agente è specificamente indicato per la vegetazione delle foreste e viene applicato puro con una concentrazione di circa 28 litri per ettaro, quantità approssimativamente 25 volte superiore a quella necessaria per applicazioni domestiche. Dopo una settimana o più dalla sua applicazione le foglie cadono dalla massima parte degli alberi. L'Arancione è stato l'agente defoliante più largamente usato in Vietnam, ma recentemente il suo impiego è stato sospeso per il fatto che il 2,4,5-T può provocare imperfezioni fisiche nelle nascite umane.

L'agente Bianco è una miscela nel rapporto 4 a 1 dei sali triisopropanolamminici del 2,4-D e dell'acido 4-ammino-3, 4, 6-tricloropicolinico. Questo ultimo componente è noto come picloram. L'agente viene spruzzato sotto forma di soluzione acquosa e con una concentrazione di circa 7,7 chilogrammi di erbicida per ettaro. Esso viene in genere adoperato con la stessa funzione dell'Arancione, per quanto la sua volatilità alquanto inferiore lo renda preferibile per operazioni nelle quali il fatto che l'agente possa venire trasportato altrove dal vento sia pericoloso, come in vicinanza di piantagioni di gomma. L'estrema resistenza del picloram alla degradazione biologica è stato un fattore che ne ha limitato l'impiego.

L'agente Blu è una soluzione acquosa di sodio dimetilarsinato, applicato alla concentrazione di circa 9 chilogrammi per ettaro; viene adoperato soprattutto contro il riso.

Qualsiasi valutazione sull'efficacia militare degli erbicidi è resa particolarmente difficile dal fatto che essi esercitano i loro effetti sul nemico solo indirettamente e dopo un notevole intervallo di tempo. Certamente la più dubbia forma di guerra a base di antipiante da un punto di vista militare è la distruzione dei raccolti. In genere non è possibile distinguere raccolti destinati ai civili da quelli che devono essere consumati dai soldati. Infatti, sia il senso comune che l'esperienza di diverse guerre dimostrano che quando le riserve di cibo vengono limitate sono i civili e non i soldati che riducono le razioni, e sono soprattutto i bambini che subiscono le conseguenze della malnutrizione.

Esistono informazioni quantitative molto limitate sulla utilità militare complessiva degli erbicidi adoperati per migliorare la visibilità. Non vi sono dubbi che le foglie cadano e quindi che la visibilità venga migliorata. Di conseguenza un nemico in genere deciderà di evitare tali aree. Tuttavia un comandante nemico non ritirerà i suoi uomini dall'azione: li impiegherà semplicemente altrove. Così per ogni soldato che guadagna in sicurezza per il fatto che la visibilità è stata migliorata può esserci un altro soldato o un civile che ci guadagnano del fuoco da parte del soldato nemico impiegato altrove. Ovviamente lo scambio non è esattamente pari, ma tale effetto fa sì che l'apparente efficacia militare degli erbicidi verrà in realtà esagerata.

In talune situazioni esistono certamente alternative ai defolianti chimici. Nel Vietnam sono stati adoperati per pulire più di 200 000 ettari degli aratri giganti, che spesso lavorano sotto la protezione di scorte militari. Quando è possibile, questo metodo è più efficace dello spruzzamento aereo dei defolianti chimici in quanto oltre che foglie vengono così rimossi anche rami e tronchi.

Allo scopo di definire i problemi ho tentato di mantenere la discussione su di un piano concreto. È facile, comunque, farsi prendere dall'emozione sulla questione delle armi chimiche e biologiche.

I sentimenti violenti provocati da tali armi sono la causa del fatto che queste storicamente siano state oggetto di sforzi speciali per il controllo degli armamenti. Il più importante tentativo internazionale di proibire l'uso di tali armi chimiche e biologiche è il Protocollo di Ginevra del 1925. Esso proibisce: 1) « l'uso in guerra di gas asfissianti, velenosi, o altri gas, e di tutti i liquidi, materiali o mezzi analoghi » e: 2) « l'uso di metodi batteriologici di guerra ». L'accordo non proibisce ricerca, sviluppo, prova o produzione di armi a gas o a germi. Non proibisce l'uso di tali armi come rappresaglia contro un nemico che le abbia usate per primo; non proibisce l'uso in guerra di fumi non tossici usati per l'occultamento o di lanciafiamme, napalm o altre armi incendiarie; non proibisce l'uso di gas per il controllo di tumulti o di altri metodi adottati dalle varie polizie. La formulazione del Protocollo di Ginevra deriva dai trattati di pace della I guerra mondiale, i quali consideravano la guerra di gas come se fosse già proibita, e proibivano specificamente la fabbricazione e l'importazione dei gas bellici dalla Germania e dai suoi alleati di guerra. Per iniziativa degli USA un articolo basato sui principi dei trattati di pace fu incluso nel Trattato di Washington del 1922 sui sottomarini e i gas nocivi. Sotto la spinta del presidente Harding, del segretario di stato Charles Evans Hughes, del senatore Elihu Root e di un comitato consultivo di eminenti cittadini nominato dal presidente, il Trattato di Washington fu approvato dal Senato senza nessun voto contrario. Sebbene ratificato dagli USA, dalla Gran Bretagna, dall'Italia e dal Giappone, questo trattato non fu mai applicato in quanto la Francia, di cui si chiedeva la ratifica, non era d'accordo sulle clausole riguardanti i sottomarini.

In seguito gli USA sollecitarono di

nuovo una proibizione della guerra di gas nel 1925 alla Conferenza di Ginevra sulla limitazione degli armamenti, proponendo sui gas una formulazione essenzialmente identica a quella del Trattato di Washington. La posizione fu estesa fino a comprendere i « metodi batteriologici di guerra » per suggerimento della Polonia. Il trattato che ne risultò – il Protocollo di Ginevra per la proibizione dell'uso in guerra di gas asfissianti, velenosi o altri gas, e dei metodi batteriologici di guerra – fu firmato dai rappresentanti di 38 nazioni il 17 giugno 1925.

L'amministrazione Coolidge e i vari sostenitori dell'accordo sembra che dessero per scontato che il Senato avrebbe dato la sua approvazione in modo altrettanto rapido del Trattato di Washington. Quasi nulla fu fatto per preparare la situazione per la ratifica o per mobilitare l'opinione pubblica. Nel frattempo l'Ufficio per la guerra chimica dell'esercito, l'American Legion, la Società chimica americana e alcune sezioni dell'industria chimica organizzarono l'opposizione. Le argomentazioni sollevate contro la ratifica erano che l'accordo sarebbe stato ignorato in tempo di guerra e che i gas velenosi erano più « umani » delle bombe e delle pallottole. L'accordo fu discusso al Senato ma non fu votato – evidentemente perché il capo della maggioranza non disponeva dei voti sufficienti. Esso rimase iscritto in ruolo presso il Comitato per gli affari esteri fino al 1947, quando il presidente Truman lo ritirò assieme a diversi altri trattati in sospeso da lungo tempo.

Nel 1939 il Protocollo di Ginevra comprese tutte le maggiori potenze europee. Allo scoppiare della II guerra mondiale l'Inghilterra, la Francia e la Germania si scambiarono assicurazioni che avrebbero rispettato l'accordo. Nel 1943 il presidente Roosevelt dichiarò che la guerra di gas era « bandita secondo l'opinione comune dell'umanità civile » e che « in nessun caso noi faremo ricorso all'uso di tali armi, a meno che non vengano adoperate prima dai nostri nemici ». Si pensa che il Giappone abbia usato i gas contro la Cina prima dell'ingresso in guerra degli USA, ma per il resto i gas non furono usati durante la II guerra mondiale. La minaccia di ritorsione ne costituiva il freno, che veniva rafforzato dalla generale avversione per gas e germi e dallo scetticismo dei militari circa la loro utilità; era però il Protocollo che poneva gas e germi in una categoria diversa e forniva un chiaro punto di riferimento sul quale i belligeranti potevano basare la loro condotta.

Sin dalla II guerra mondiale gli USA hanno dichiarato in diverse occasioni il loro appoggio al principio, contenuto nel Protocollo di Ginevra, di non usare per primi queste armi. Quando fu chiesto al presidente Eisenhower in una conferenza stampa se egli intendesse mutare la nostra politica del non-uso per primi, egli disse: « Nessuno mi ha ufficialmente consigliato di farlo, e, per quanto riguarda il mio istinto, ritengo che non bisogna dare inizio per primi a questa cosa ». Durante l'amministrazione del presidente Johnson gli USA appoggiarono la risoluzione approvata nel 1966 e nel 1968 dall'Assemblea generale delle Nazioni Unite, che richiamava a una « stretta osservanza da parte di tutti gli stati dei principi e degli obiettivi del Protocollo » e « invitando tutte le nazioni che non lo hanno fatto a sottoscrivere il Protocollo ». Tuttavia alcune dichiarazioni hanno fatto apparire ambigua la posizione degli Stati Uniti. Per esempio, il manuale dell'esercito americano *The Law of Land Warfare*, pubblicato fino agli anni 50, afferma che il Protocollo di Ginevra « non obbliga il nostro paese ».

Complessivamente 84 nazioni fanno ora parte del Protocollo di Ginevra, 16 delle quali lo hanno ratificato subito dopo la risoluzione del 1966 dell'ONU. Gli aderenti includono tutti i membri della NATO a eccezione degli USA, tutti i membri del Patto di Varsavia compresa l'URSS e tutte le potenze nucleari (eccetto gli USA), compresa la Repubblica Popolare Cinese. Tutte le maggiori nazioni industriali, eccetto il Giappone e gli Stati Uniti, ne fanno parte. L'accordo è stato firmato ma non ancora ratificato da USA, Brasile, El Salvador, Giappone, Nicaragua e Uruguay.

La questione se l'accordo proibisca o no l'uso di agenti aggressivi e defolianti chimici è ancora oggetto di qualche divergenza. In un primo momento quando le nazioni venivano sollecitate perché esprimessero i loro punti di vista sullo stato dei gas lacrimogeni (nella Commissione della Società delle Nazioni, nel 1930), il Canada, la Cina, la Francia, l'URSS e diverse altre nazioni si dichiararono d'accordo con la posizione dichiarata dell'Inghilterra che cioè « l'uso in guerra di "altri" gas, compresi quelli lacrimogeni era proibito ». Nessuna delle nazioni che allora facevano parte del Protocollo fece obiezioni a questa presa di posizione. Il delegato degli USA, tuttavia, manifestò esitazioni su qualsiasi impegno a evitare in guerra l'uso di agenti che fossero adoperati in periodo di pace dalla polizia nazionale e il cui uso in combattimento sarebbe stato « più chiaramente umano di quelli cui precedentemente si era costretti a ricorrere ». Due anni dopo la Conferenza per il disarmo indetta dalla Società delle Nazioni raccomandò unanimemente che fosse proibito l'uso in guerra di tutti i gas, compresi quelli lacrimogeni. Tale punto di vista fu accettato dagli USA con l'intesa che non fosse esteso all'uso dei gas lacrimogeni adoperati per funzioni di polizia. Le discussioni non si svolgevano in funzione del Protocollo di Ginevra ma di un trattato globale sul disarmo, un tentativo che venne frustrato dall'approssimarsi della II guerra mondiale. La questione non tornò a galla che nel 1965, quando vennero per la prima volta sollevati interrogativi sull'impiego di CS e di erbicidi da parte americana nel Vietnam.

Rispondendo a tali interrogativi nel 1966 il rappresentante americano all'Assemblea generale affermò che l'accordo non proibisce « l'uso in combattimento contro un nemico, per scopi umanitari, di agenti che tutti i governi del mondo usano comunemente per controllare i tumulti del proprio popolo » o di sostanze chimiche usate per controllare la « vegetazione indesiderata » per scopi agricoli o altri scopi pacifici. In quell'epoca questi punti di vista incontrarono l'opposizione soprattutto dell'URSS e dei suoi alleati. Tuttavia, man mano che le dimensioni delle operazioni chimiche nel Vietnam si accrescevano, la posizione degli USA era sottoposta a un attacco sempre crescente. L'estate scorsa il Segretario generale U Thant spinse i membri dell'ONU a « fare una chiara dichiarazione » che l'accordo proibisce l'uso in guerra di tutti gli agenti chimici, compresi i gas lacrimogeni. Lo scorso inverno l'India, il Messico, il Pakistan, la Svezia e altre 17 nazioni proposero all'Assemblea generale una risoluzione che sosteneva che il Protocollo di Ginevra proibisce l'uso bellico di tutti gli agenti chimici rivolti contro l'uomo, gli animali o le piante. La risoluzione fu approvata con 80 voti a favore e tre contrari, e con 36 astensioni. Contro la risoluzione votarono Portogallo, Australia e Stati Uniti.

In seguito il governo inglese ha dichiarato che il CS non rientra nel Protocollo di Ginevra, sebbene l'Inghilterra ammetta ancora che i gas lacrimogeni noti nel 1930 siano proibiti. La giustificazione inglese è che il CS è meno tossico dei vecchi gas lacrimogeni, ma si ritiene generalmente che il motivo reale alla base di tale dichiarazione è di evitare il rischio che l'uso del CS per il controllo delle sommosse scoppiate nell'Ulster sia accusato come illegale. Se è così, ciò riflette una confusione sull'ambito del Protocollo di Ginevra, il quale secondo le sue clausole proibisce l'uso dei gas solo in situazioni belliche. La distinzione fra l'uso di gas lacrimogeni per scopi interni di polizia e la proibizione del suo uso in guerra è stata sempre osservata senza difficoltà da diverse nazioni sin dalla II guerra mondiale. Anche la Svezia, una nazione che ha guidato l'opposizione contro l'uso di gas lacrimogeni in guerra, ne fa uso per il controllo dei tumulti in casa. Al colmo dell'ironia, potrebbe avvenire che proprio l'uso massiccio di CS in guerra e gli eccessi che ciò genera provochino una seria opposizione anche al suo uso interno.

Dato che gli USA intendono ratificare il Protocollo di Ginevra, l'amministrazione Nixon deve prendere una decisione su una politica relativa all'uso di agenti aggressivi ed erbicidi. Non ci sono molte difficoltà a identificare i pro e i contro il fatto di mantenere la possibilità di scelta circa l'uso di tali agenti in guerra. Le difficoltà nascono quando si cerca di dare un peso preciso agli elementi a favore e contro, allo scopo di trarne delle conclusioni.

Il principale vantaggio di chi fa uso di tali agenti in guerra consiste nel fatto che in certa misura questi gli consentono di accrescere le perdite nemiche e di ridurre quelle proprie. Tali agenti, tuttavia, hanno un effetto significativo quasi solo nelle azioni belliche anti guerriglia e, nel caso particolare del CS, solo se le truppe nemiche non sono protette da maschere antigas. Perciò gli argomenti a favore del fatto di conservare queste armi diventano importanti solo nel caso e nella misura in cui una nazione preveda di essere implicata in guerre di questo tipo. In ogni caso è prevedibile che il CS perda i suoi vantaggi perché, una volta che si sarà radicata l'impressione che le forze armate di un paese faranno uso di gas, l'uso delle maschere antigas diverrà molto più diffuso in tutto il mondo. Le truppe nord-vietnamite nel Vietnam del sud sono ora largamente fornite di ottime maschere antigas di fabbricazione cinese mentre aumenta sempre di più il numero di maschere russe di tipo Shlem. Anche per dei guerriglieri il costo di queste maschere non è elevato.

La più grossa minaccia alla sicurezza derivante dal fatto di seguitare a far uso di questi agenti consiste nell'accresciuta probabilità di proliferazione di armi chimiche e annullamento dei freni posti alla guerra chimica. L'impiego da parte americana di gas aggressivi in guerra, in particolare quando ciò viene fatto in larga scala e associato alle ordinarie operazioni militari, spinge in altre nazioni i responsabili militari a fornirsi di maschere antigas, a dare inizio all'addestramento chimico e preparare i quadri di esperti chimici e a prendere in considerazione l'acquisto per le proprie truppe di armi chimiche di vario tipo, comprese quelle letali.

L'impiego militare in larga scala di sostanze chimiche defolianti pone problemi ecologici e di salute pubblica sui quali sappiamo ancora poco ma che a lunga scadenza potrebbero diventare seri. Il precedente di favorire alterazioni in larga scala dell'ambiente durante la conduzione di una guerra presenta ovvi pericoli. Inoltre, in diverse parti del mondo si potrebbe far uso di agenti chimici come mezzi di guerra economica e di affamamento contro popolazioni civili. Data la relativa facilità con cui questa tattica può essere praticata, sarebbe difficile fermarla una volta che si siano creati i precedenti.

Il problema generale di evitare le guerre chimiche e biologiche è in gran parte psicologico. Forse il problema centrale consiste nell'evitare l'applicazione della biochimica e della biologia in una direzione che aprirebbe una nuova e imprevedibile dimensione alla guerra. Se siamo in grado di difendere e rafforzare la ormai tradizionale speranza che nessun gas né germe venga usato in guerra, non si manifesterà nessuna forte pressione per la proliferazione di tali armi.

Tale aspetto psicologico del problema è stato compreso essenzialmente da tutte le nazioni, anche dagli USA, sin dalla I guerra mondiale. Recentemente, tuttavia, si è lasciato che si verificasse e si amplificasse nel Vietnam una pericolosa rottura con la tradizione. Molti di coloro che hanno studiato il problema considerano l'uso di armi chimiche in quel paese, anche se si tratta di sostanze chimiche non letali, come la maggiore e più immediata minaccia alle barriere che prevengono finora la guerra chimica.

Se saremo capaci di rinunciare alla possibilità di scelta a nostra discrezione sull'uso di agenti aggressivi e di sostanze chimiche defolianti come armi belliche, allora, a giudicare dal recente voto all'Assemblea generale, sembra probabile che tutte le nazioni di rilievo del mondo potranno essere spinte ad accordarsi. Se, invece, verrà abbandonata la legge tanto a lungo osservata, ossia « niente armi chimiche e biologiche », non esisterà più una norma unica e imparzialmente semplice sulla quale si possano basare la pratica nazionale e l'accordo internazionale.

Effetti di attacchi a base di agenti chimici. La valutazione si basa sull'ipotesi che l'attacco sia condotto da un solo aeroplano, lungo una linea di due chilometri controvento, scaricando quattro tonnellate di agente. L'uso di VX in forma di aerosol consentirebbe di spingere al massimo lo spostamento dell'agente nella direzione del vento.

Marte come appariva da una distanza di 356 000 km il 4 agosto 1969 al teleobiettivo della telecamera di *Mariner 7*. Circa 27 ore piú tardi la sonda ha ripreso porzioni della stessa faccia del pianeta a distanza inferiore; pochi giorni prima quando il *Mariner 6* aveva ripreso immagini a distanza ravvicinata il pianeta rivolgeva alla telecamera pressappoco la stessa faccia. Per questa ragione le stesse strutture visibili in questa immagine ripresa a grande distanza possono essere paragonate con i dettagli ripresi precedentemente dal *Mariner 6* e successivamente dallo stesso *Mariner 7*. Degno di nota il fatto che in tutte le immagini manca qualunque cosa possa rassomigliare al reticolato di canali che da molto tempo gli studiosi del pianeta avevano rappresentato sulle loro carte e che tuttora appare in carte recenti del pianeta (*si vedano le illustrazioni alle pag. 26 e 27*). In queste e in tutte le immagini riprese a grande distanza dai due *Mariner* appaiono chiaramente le maggiori strutture della superficie del pianeta visibili con i maggiori telescopi terrestri. Cosí la grande regione scura che si estende al di sotto dell'Equatore inizia sulla sinistra con la struttura bifida chiamata Meridiani Sinus. La piú sottile banda scura immediatamente a destra è Sabaeus Sinus che quando ruota verso il basso prende il nome di Mare Serpentis. La regione scura triangolare che si estende verso nord sopra l'equatore sul lato destro del pianeta è Syrtis Major. Al di sotto di essa, poco sopra la calotta polare, vi è il grande « deserto » circolare chiamato Hellas, che risulta privo di strutture superficiali.

La superficie di Marte

Le chiarissime immagini riprese dai Mariner 6 e 7 mostrano un pianeta, caratterizzato da strutture superficiali diverse sia da quelle della Luna sia da quelle della Terra, le cui calotte polari sono costituite da ghiaccio secco

di Robert B. Leighton

La prima rappresentazione grafica di Marte, ottenuta con l'aiuto del telescopio, mostra una sola macchia nera rotonda al centro di una sfera perfettamente distinta. Lo schizzo fu realizzato da un astronomo italiano nel 1636, Franciscus Fontana. Trenta anni piú tardi Gian Domenico Cassini osservò le calotte polari del pianeta e produsse diversi schizzi che mostrano tutta una serie di zone illuminate e zone scure sparse qua e là sulla sua superficie; dai mutamenti di posizione di tali zone egli ne ricavò che il periodo di rotazione doveva essere di 24 ore e 40 minuti, con un errore complessivo inferiore ai tre minuti.

Oltre alle calotte polari, i disegni di Cassini riportano almeno due strutture circolari caratterizzate da luminosità particolarmente intensa che possono corrispondere, rispettivamente alle regioni ora note come Nix Olympica e deserto Elysium.

Tra il 28 luglio e il 5 agosto dello scorso anno la Nix Olympica, il deserto Elysium e decine di altre strutture di Marte, disegnate, ridisegnate e interpretate piú volte nei tre secoli trascorsi da quando Fontana e Cassini hanno eseguito le loro prime osservazioni, sono state riprese attraverso gli obbiettivi delle telecamere portate fin quasi sul pianeta dai *Mariner 6* e *Mariner 7*: le due sonde hanno inviato sulla Terra 202 immagini complete riprese a distanze variabili tra un massimo di 1 716 000 e un minimo di 3500 chilometri dalla superficie del pianeta. Altre 1177 fotografie, ciascuna contenente un settimo delle informazioni contenute nelle immagini complete, sono state riprese a distanze sia superiori che inferiori a 1 800 000 chilometri: anche se non sono immagini molto belle a vedersi, costituiscono pur sempre una ricca fonte di informazioni.

Ciascuna delle 202 immagini complete contiene piú unità di informazione dell'intera serie delle 22 fotografie riprese nel luglio 1965 dalla sonda *Mariner 4*, la prima che abbia fotografato un altro pianeta. Il *Mariner 4* non era progettata per fotografare l'intero disco di Marte e le fotografie ravvicinate che essa ci ha fornito, pur notevolmente interessanti, non sono servite a gettare nuova luce sulle strutture a grande scala visibili da Terra: in esse si poteva osservare una superficie assai ricca di crateri che assomigliava molto a quella della Luna. Tale spettacolo sarebbe stato una vera sorpresa per gli studiosi di Marte del XIX secolo che pensavano al pianeta rosso come a un gemello della Terra.

La nuova sorpresa riservata dalle fotografie dei *Mariner 6* e *7* rispetto a quelle del *Mariner 4* è stata quella di mostrare che Marte non è una Luna di grandi dimensioni ma che invece è un corpo celeste la cui superficie è caratterizzata da strutture sconosciute in tutto il resto del sistema solare: oltre alle zone assai ricche di crateri rivelateci da *Mariner 4*, vi sono su Marte alcune grandi aree, come il deserto Hellas, che ne sono praticamente prive. Quale fenomeno può avere provocato la scomparsa dei crateri che quasi sicuramente esistevano una volta in quelle stesse zone? Sulle stesse fotografie è stato riconosciuto un terzo tipo di struttura superficiale che conferisce un aspetto caotico alle zone in cui si sviluppa. Anche in esse i crateri sono stati per la maggior parte distrutti, ma, nel contempo, deve essere accaduto qualcosa che ha determinato la formazione di una superficie a dorsali ad andamento caotico che non trova riscontro in nessuna struttura di analoghe dimensioni né sulla Luna né sulla Terra.

Mariner 6 e *Mariner 7* erano inoltre equipaggiate di spettrometri che avevano lo scopo di studiare la composizione dell'atmosfera di Marte nelle regioni infrarossa e ultravioletta dello spettro, insieme con un radiometro all'infrarosso per misurare la temperatura superficiale del pianeta. Altre informazioni sulla composizione dell'atmosfera di Marte sono state ottenute studiando le piccole variazioni di frequenza dei segnali radio provenienti dalle sonde, nei momenti in cui esse scomparivano dietro il pianeta e, qualche minuto dopo quando uscivano nuovamente dall'occultazione. Un gruppo finale di esperimenti condotti a termine tramite il *Mariner 6* e *7* utilizza i dati relativi alla rotta seguita dalle sonde per stabilire le esatte dimensioni del sistema solare e le orbite di Marte, della Terra e dei pianeti adiacenti. In questo articolo si tratterà soltanto delle fotografie di Marte e della loro interpretazione: per la messa a punto di questo articolo ci si è giovati anche della collaborazione di Norman H. Horowitz, Bruce C. Murray e Robert P. Sharp del California Institute of Technology, Alan G. Herriman e Andrew T. Young del Jet Propulsion Laboratory, Bradford A. Smith della New Mexico State University, Merton E. Davies della Rand Corporation e Conway B. Leovy della University of Washington.

Nel dicembre 1965, cinque mesi dopo il fortunato lancio del *Mariner 4*, il Jet Propulsion Laboratory diretto per la NASA (National Aeronautics and Space Administration) dal Cal Tech, fu autorizzato a progettare i *Mariner 6* e *7* e a sovrintendere alle loro missioni verso Marte. La prima sonda spaziale era stata messa in orbita da un missile a due stadi Atlas-Agena, capace di far giungere in prossimità di Marte una massa totale di circa 260 chilogrammi: la telecamera pesava 15 chilogrammi e l'equipaggiamento scientifico rimanente 25 chilogrammi. *Ma-*

L'aspetto della faccia di Marte centrata circa a 120° di longitudine est, in una foto da un telescopio terrestre (*qui sopra*) e in tre immagini (*su questa e sulla pagina a fronte*) riprese dai *Mariner 6* e *7* mentre si avvicinavano al pianeta. *Mariner 7* ha ripreso la prima e la terza immagine a distanze di 848 226 e 255 278 chilometri. *Mariner 6* ha ripreso la seconda immagine a una distanza di 328 075 chilometri. Le longitudini centrali nelle immagini dei *Mariner* sono rispettivamente 99°, 105° e 119° di longitudine est. La massa scura che appare all'estrema sinistra è Syrtis Major. In alto a destra in tutte le foto il grande deserto circolare chiamato Elysium; altrettanto ben visibili anche la struttura Trivium Charontis e il «canale» Cerberus.

riner 6 e *Mariner 7* sono stati messi in orbita da un vettore costituito dalla combinazione di un Atlas e di un Centaur, in cui il secondo stadio (il Centaur) usa come carburante idrogeno liquido. È stato cosí possibile accrescere il peso complessivo della sonda fino a 380 chilogrammi, cosicché il peso complessivo della strumentazione scientifica è stato all'incirca triplicato: da 40 a 115 kg. Ciò ha reso possibile missioni con due telecamere, una per coperture grandangolari, l'altra per riprese con teleobiettivo.

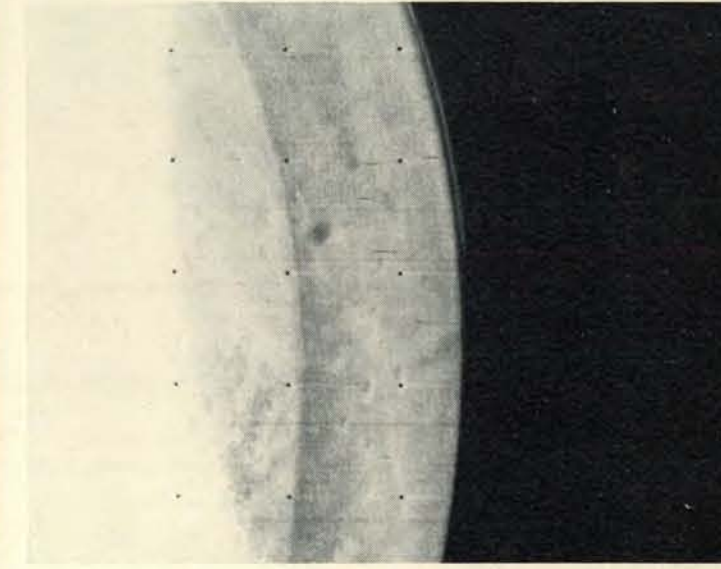
La telecamera del *Mariner 4* era dotata di una lunghezza focale di 305 millimetri e produceva un'immagine all'ingresso del tubo da ripresa vidicon delle dimensioni di 5,5 per 5,5 millimetri. Il vidicon trasformava ciascuna immagine in una trama costituita da 200 linee, ciascuna formata da 200 elementi, detti *pixel*, per un totale di 40 000 *pixel* per immagine. Ciascun *pixel* era previsto per 64 livelli di luminosità, che richiedevano sei *bit* di informazione per *pixel* ($2^6 = 64$). Il numero totale di *bit* era dunque di sei volte 40 000, ossia 240 000. Nel momento in cui si trovava piú vicina a Marte (9850 km) la telecamera del *Mariner 4* copriva una superficie di circa 300 km di lato ed era in grado di rivelare crateri di poco meno di cinque chilometri di diametro.

Nei *Mariner 6* e *7* la telecamera a grandangolo (telecamera A) aveva una lunghezza focale di 52 mm mentre l'altra (telecamera B) aveva una lunghezza focale di 508 millimetri. Nel momento in cui le sonde si sono trovate piú vicine a Marte (distanza di 3500 km) la telecamera B copriva una superficie di circa 72 $\times$ 84 km con una capacità risolutiva che le permetteva di rilevare crateri di poco meno di 300 metri di diametro. Nel contempo la telecamera A copriva una zona rettangolare i cui lati erano dieci volte piú lunghi per un'area cento volte piú grande.

All'ingresso del tubo da ripresa vidicon si formava un'immagine rettangolare i cui lati erano 9,6 e 12,3 millimetri. Tale immagine veniva convertita in un'immagine televisiva costituita da 704 linee, ognuna formata da 935 *pixel*, per un totale di 658 240 *pixel* (la televisione italiana usa immagini costituite da 625 linee, mentre quella commerciale americana immagini da 525 linee ognuna costituita da 400 elementi). Ciascun *pixel* era previsto per 256 livelli di luminosità, richiedendo otto *bit* d'informazione ($2^8 = 256$). Cosí ogni immagine ripresa dai *Mariner 6* e *7* conteneva cinque milioni di *bit* d'informazione, contro i 240 000 del *Mariner 4*. (La codificazione in forma numerica di una immagine televisiva normale richiederebbe circa un milione di *bit*).

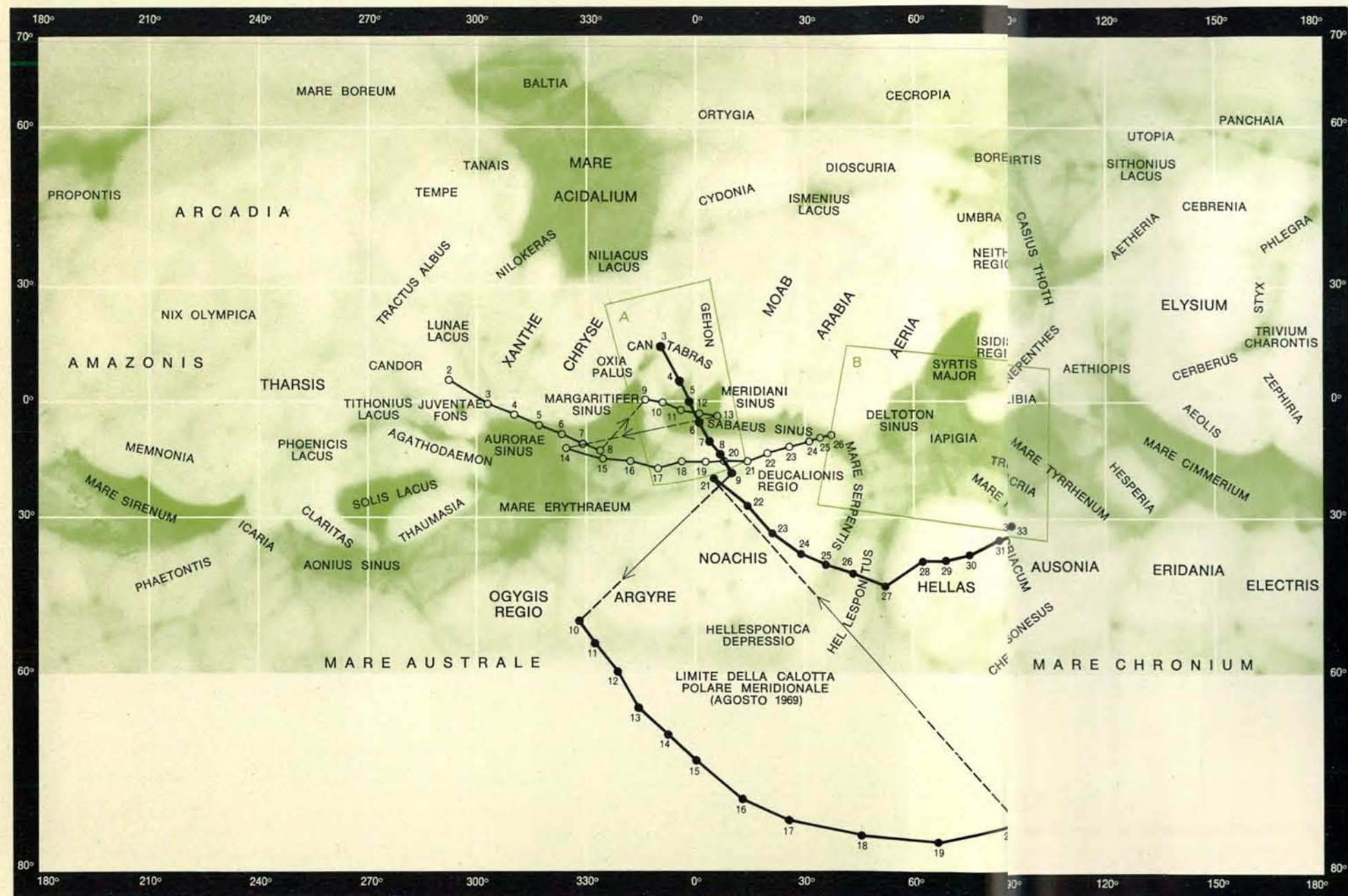
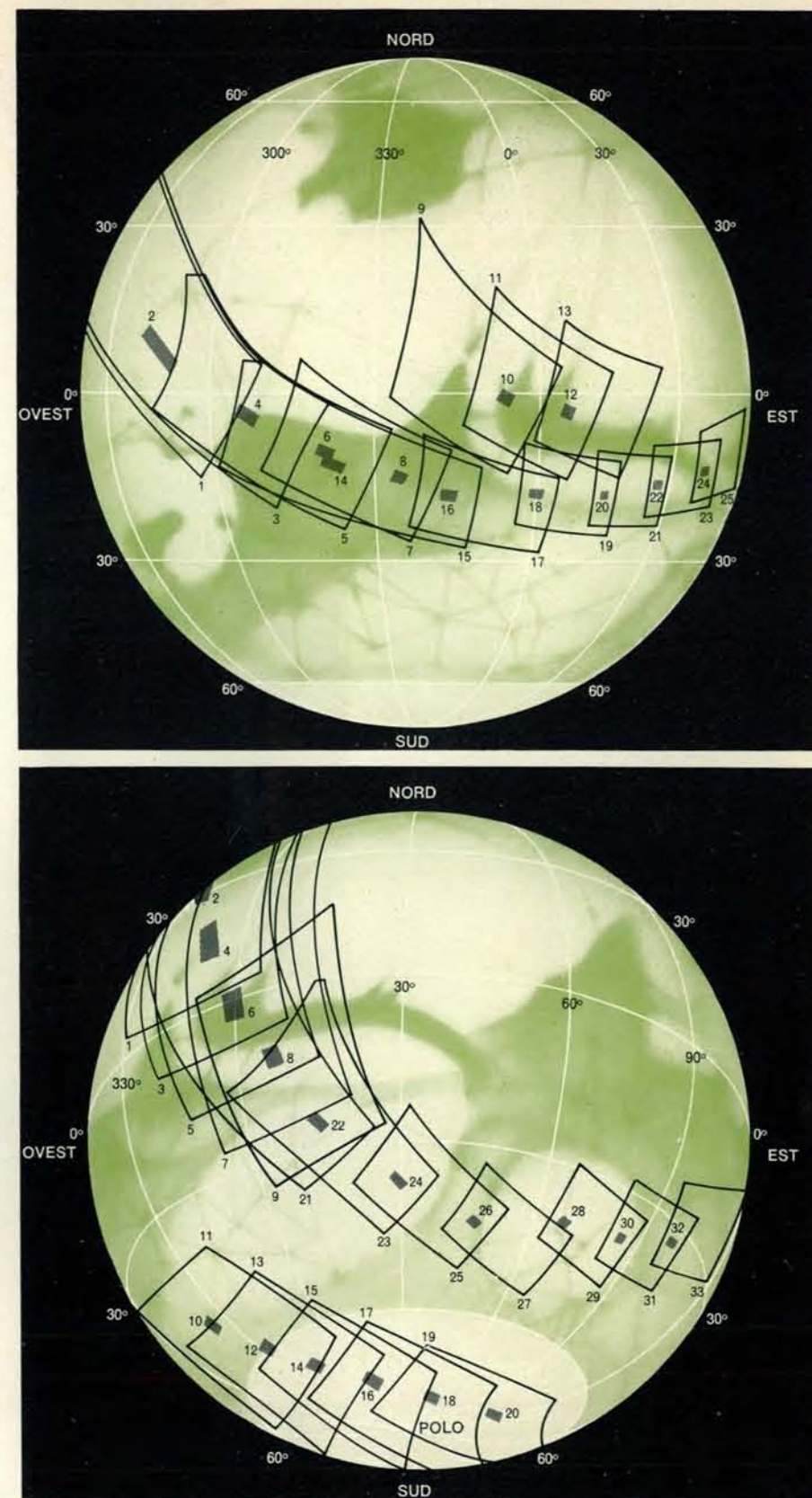
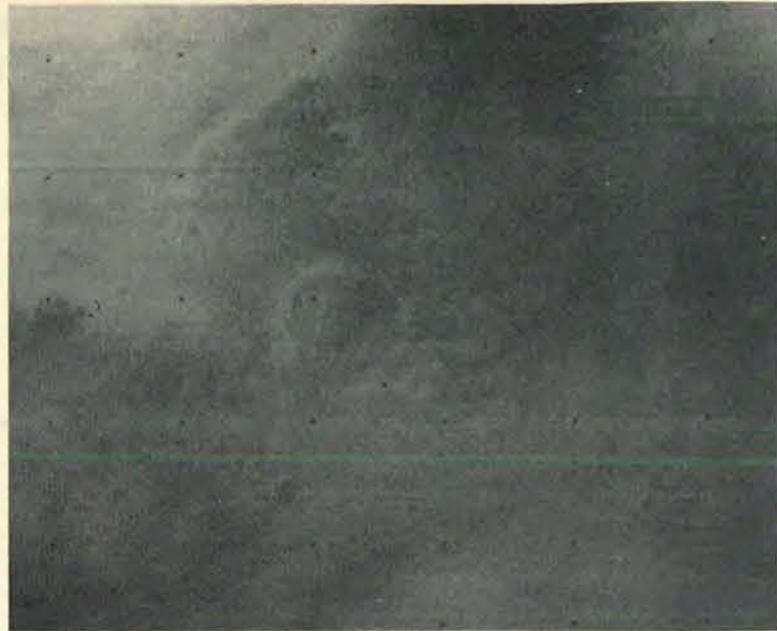
A causa di diverse limitazioni che spiegherò brevemente, le immagini non potevano essere ritrasmesse a Terra in tempo reale come fa di norma la televisione ma dovevano essere registrate su nastro magnetico. Il nastro usato per *Mariner 4* aveva una capacità di accumulo di informazioni di circa cinque milioni di *bit*, sufficienti per 22 immagini. Queste sono state successivamente ritrasmesse a Terra alla velocità di otto *bit* e mezzo per secondo cosicché la trasmissione a Terra dell'intera serie di immagini ha richiesto piú di otto ore. Per *Mariner 6* e *7* abbiamo scelto un sistema che richiede l'uso di due nastri: uno numerico con una capacità di tredici milioni di *bit*, e uno analogico con una capacità effettiva di 120 milioni di *bit*. Entrambi i nastri avevano all'incirca la stessa lunghezza (110 metri), lo stesso numero di tracce (quattro) e la stessa velocità di scorrimento (trenta centimetri al secondo). La ragione della grande differenza nella capacità effettiva è che il nastro numerico è usato per immagazzinare solo un *bit* di informazione (0 o 1) per ogni unità magnetica: quest'ultima può essere magnetizzata oppure no e non esiste quindi ambiguità. In una registrazione analogica invece la stessa unità magnetica può essere magnetizzata piú o meno intensamente a seconda dell'intensità del segnale ricevuto. In seguito il segnale analogico può essere trasformato in forma numerica ma il risultato è meno sicuro di quanto non sarebbe stato se l'informazione fosse stata registrata direttamente in forma numerica.

Sulle sonde *Mariner 6* e *Mariner 7* il segnale analogico proveniente dal tubo da ripresa vidicon è stato trattato in

Il bagliore pomeridiano di certe regioni di Marte, i cui mutamenti erano già stati osservati anche da Terra, è evidente in diverse delle immagini riprese dai *Mariner* da grande distanza. Queste due immagini (n. 76 e 78) cosí come la n. 79 che si trova sulla pagina a fronte mostrano come si modifica tale bagliore nella regione Tharsis-Candor nel primo pomeriggio. Nell'immagine n. 76 (longitudine 199° est) il bagliore non è visibile; le immagini successive, riprese 95 e 114 minuti piú tardi mostrano una zona luminosa al margine destro del pianeta. Un sottile strato nebuloso lungo i contorni di Marte è visibile in diverse foto a distanza ravvicinata riprese da *Mariner 7*. La banda scura sulla superficie del pianeta è solo un artificio.

due modi. Una parte di esso è stato tradotto in un codice a otto *bit*: il livello di luminosità zero è fatto corrispondere a 00000000 mentre il massimo corrispondeva a 11111111. I primi due segnali numerici sono stati mediati su diverse linee e trasmessi a Terra in tempo reale come parte del flusso di dati telemetrici. I sei *bit* restanti di ogni settimo *pixel* sono stati registrati dal nastro per la registrazione numerica e quindi ritrasmessi a Terra producendo le 1177 immagini ciascuna delle quali completa solo per un settimo.

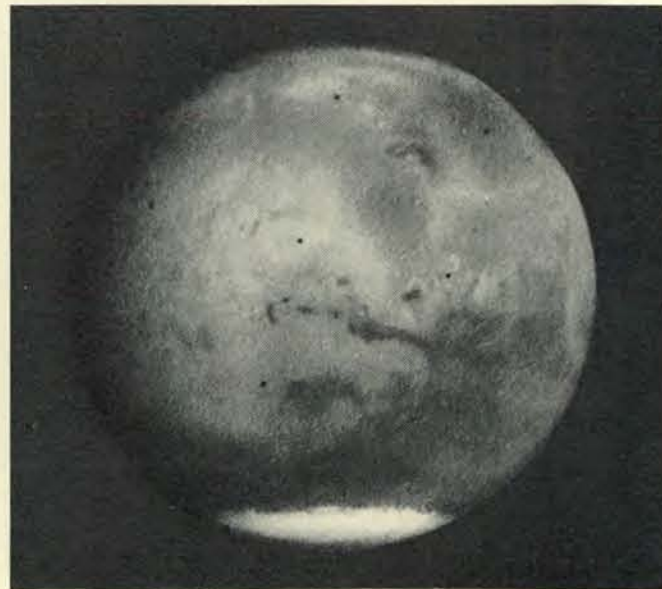
Simultaneamente un altro campione del segnale analogico del vidicon è stato passato sul nastro a registrazione analogica dopo essere stato modificato in due modi. Allo scopo di intensificare la visibilità dei particolari il segnale analogico è stato automaticamente controllato in modo che il suo valore medio fosse approssimativamente costante, con una tecnica chiamata *automatic gain control*. Inoltre il segnale è stato immesso in un circuito che ne ha aumentato il contrasto locale secondo un fattore tre. L'effetto finale di questo trattamento è stato quello di fornire un segnale meglio classificabile quando successivamente esso è stato convertito in un segnale numerico a sei *bit* per essere ritrasmesso a Terra.

Piú tardi i primi due *bit* del segnale numerico che erano stati inizialmente mediati su diverse linee e quindi ritrasmessi a Terra, sono stati immessi in computer e riuniti con i dati provenienti dalla registrazione analogica a dare un'immagine completa a otto *bit*, cioè a 256 livelli.

Se la luminosità dei successivi *pixel* fosse variata in modo casuale da 0 a 255, lo schema non avrebbe funzionato.

Una carta di Marte che raccoglie le osservazioni telescopiche eseguite durante oltre tre secoli, nonostante che difficilmente due osservatori diversi si siano mai trovati d'accordo sul reale aspetto della superficie del pianeta. Questa proiezione di Mercatore della superficie di Marte è basata su un resoconto preparato pochi anni fa dall'Aeronautical Chart and Information Center della US Air Force. Anch'essa mostra, secondo la tradizione, la superficie del pianeta attraversata da un reticolo di canali. Tuttavia la carta riporta anche i crateri fotografati da *Mariner 4* nel 1965: essi si estendono da Trivium Charontis (*estrema destra*) fino a Phaetontis (*estrema sinistra*). È affascinante paragonare le strutture riportate sulla carta con i complessi dettagli visibili sulle fotografie inviate a Terra da *Mariner 6* e *7*. Le linee colorate comprendono zone riprese in una delle ultime immagini da grande distanza (B) e in una delle prime immagini riprese da distanza ravvicinata (A). Le due immagini sono riprodotte a pag. 28 in alto. *Mariner 6* ha ripreso 26 immagini a distanza ravvicinata, *Mariner 7* ne ha riprese 33: i circoletti vuoti numerati da 2 a 26 identificano i centri delle immagini riprese da *Mariner 6*; il centro della prima di esse non cade sul pianeta; i punti pieni numerati da 3 a 33 identificano i centri delle immagini riprese da *Mariner 7*. Durante ciascuna sequenza la piattaforma che reggeva le telecamere è stata ruotata per riprendere una serie di immagini su regioni che venivano a trovarsi a lato della rotta di volo principale; le immagini contraddistinte da numero d'ordine dispari sono state riprese con obbiettivo grandangolare.

La copertura fotografica di Marte è riportata su due globi: su quello in alto le immagini a distanza ravvicinata riprese da *Mariner 6*, su quello sotto quelle riprese da *Mariner 7*. Il globo superiore è centrato su una longitudine di 335° est; quello sotto è stato ruotato verso sinistra cosicché il meridiano centrale è 30° est e inclinato in modo da mostrare la calotta polare meridionale. Le immagini realizzate con obbiettivo grandangolare che sono contraddistinte da numero dispari, sono state riprese con filtri nella successione blu, verde, rosso, verde e cosí via; le altre immagini, indicate con numeri pari sono state riprese con teleobiettivo e con filtro giallo. La risoluzione delle immagini a grandangolo è di circa 100 volte migliore di quella ottenuta da Terra; la risoluzione delle altre riprese con teleobiettivo è all'incirca 1000 volte migliore.

Queste due immagini riprese da *Mariner 7* comprendono zone di particolare interesse. Le strutture superficiali possono essere identificate facendo riferimento alle zone A e B della carta riportata alle pagine 26 e 27. L'immagine a destra (B) è una delle ultime riprese da grande distanza (n. 92): questo tipo di immagini è stata trasmessa senza sottoporre il segnale ripreso dalla telecamera ad alcun procedimento per aumentarne il contrasto che così è stato restituito a Terra esattamente come registrato. Non è tale il caso delle immagini riprese a distanza ravvicinata, come quella sopra a sinistra (A). La zona scura visibile sullo sfondo di A è il Meridiani Sinus. I diversi crateri sparsi sopra coincidono con la posizione del «canale» Cantabras. La fotografia B risulta centrata su Iapygia e mostra una parte del Syrtis Major e il margine settentrionale di Hellas.

Tuttavia noi sapevamo dal *Mariner 4* che la superficie di Marte non ci avrebbe riservato questa sorpresa e che sarebbe stata caratterizzata da un contrasto piuttosto basso.

Una delle piú impressionanti differenze tra *Mariner 4* e i *Mariner 6* e *7* è stata la quantità di dati ritrasmessi per unità di tempo: 16 200 *bit* al secondo contro gli otto *bit* e mezzo al secondo della missione precedente. Molti fattori hanno contribuito a questo progresso, non ultimo il fatto che Marte era, nel 1969 assai piú vicino alla Terra di quanto non lo fosse nel 1965: era a soli 100 milioni di chilometri invece che a 210 milioni. Inoltre le radio trasmittenti a bordo dei *Mariner 6* e *7* raggiungevano i 20 watt, cioè il doppio della potenza disponibile sul *Mariner 4*. L'antenna ricevente di Goldstone (California) aveva un'apertura di 70 metri garantendo una superficie di ricezione sette volte maggiore di quella di 26 metri disponibile nel 1965.

Da ultimo, tra il 1965 e il 1969, ci sono stati progressi considerevoli nell'elettronica, fra cui la introduzione di un nuovo sistema di ricerca dell'errore: tutto ciò ha consentito di raggiungere la velocità di trasmissione dei dati di 16 200 *bit* al secondo.

Malgrado l'elevata velocità di trasmissione, essa non ha potuto adeguarsi alla velocità con cui l'informazione veniva raccolta dalle telecamere dei *Mariner*: 100 000 *bit* al secondo. Per questo motivo i dati hanno dovuto essere registrati e quindi ritrasmessi a Terra. È bene notare inoltre che la velocità di 16 200 *bit* al secondo ha potuto essere sviluppata soltanto in quelle sei ore al giorno in cui il segnale veniva raccolto dalla grandissima antenna di Goldstone: durante il resto della giornata, quando hanno dovuto essere usate altre stazioni riceventi in altre parti del mondo, la velocità di ritrasmissione ha dovuto essere abbassata considerevolmente (270 *bit* al secondo).

Nel progetto meccanico *Mariner 6* e *Mariner 7* possono esser considerati gemelli dei precedenti *Mariner* che, con tanta precisione sono giunti in prossimità sia di Marte che di Venere. Tuttavia la piú importante innovazione deve essere considerata l'introduzione di un nuovo computer (CC&S): il sistema elettronico che ritiene nella propria memoria il programma della missione. Il CC&S è stato programmato prima del lancio secondo una missione standard e una missione di riserva. La missione standard, che poteva essere portata a termine soltanto su comando esplicito da Terra, avrebbe dato il massimo di copertura e ritrasmissione di immagini. La missione di riserva che si sarebbe innescata spontaneamente, avrebbe restituito soltanto otto immagini del pianeta nel suo insieme e le avrebbe ritrasmesse a Terra a 270 *bit* al secondo. Un'importante caratteristica del CC&S era di poter essere riprogrammato in volo secondo istruzioni inviate da Terra. Era stato infatti anticipato che il programma del *Mariner 7* avrebbe potuto essere riveduto sulla base dei dati raccolti dal *Mariner 6*, e così è stato.

Per le due sonde erano state scelte rotte differenti. Per il *Mariner 6* era stata scelta la data di arrivo del 31 luglio e un punto di mira che avrebbe consentito di ottenere immagini lungo una fascia compresa tra l'equatore e 20° di latitudine sud. Sulla zona compresa entro ogni immagine grandangolare doveva essere ripresa un'immagine a copertura ridotta. La serie di immagini parzialmente sovrapposte avrebbe coperto almeno mezzo giro attorno al pianeta. A mezza strada la piattaforma sulla quale era montata la telecamera sarebbe stata ruotata verso nord per inquadrare la struttura prominente e oscura nota come Meridiani Sinus. La copertura principale avrebbe invece compreso un certo numero di regioni ben studiate e due strutture da lungo tempo note ai cartografi di Marte come «oasi» (Juventae Fons e Oxia Palus) e una regione a luminosità variabile, la Deucalionis Regio (*si vedano le illustrazioni alle pagine 26 e 27*).

Cinque giorni piú tardi *Mariner 7* si sarebbe avvicinato al pianeta seguendo una rotta piú meridionale per riprendere una serie di immagini con un certo grado di sovrapposizione reciproca, serie che sarebbe iniziata poco a nord dell'equatore per procedere poi verso sud-est. Questa sequenza si sarebbe poi interrotta a mezza strada per dare la possibilità di riprendere una serie di immagini del margine della calotta polare meridionale e del polo stesso. In questa sequenza sarebbe stato incluso anche Hellespontus, che appare meno luminoso non appena la calotta polare inizia a recedere, nonché il deserto Hellas. Originariamente era stato previsto di riprendere soltanto tre immagini grandangolari della regione polare meridionale.

Dopo aver osservato nelle immagini da grande distanza riprese da *Mariner 6* alcune caratteristiche atmosferiche inattese in questa regione, abbiamo richiesto al controllo della missione di riprogrammare il CC&S al fine di ottenere cinque immagini grandangolari della regione polare meridionale.

Nel riprendere le immagini ravvicinate, le telecamere a grandangolo e con teleobiettivo hanno funzionato alternativamente, con un intervallo di 42 secondi tra un'esposizione e la successiva. La telecamera a grandangolo era equipaggiata con filtri rosso verde e blu, che erano collocati in quattro cavità dietro l'otturatore. Sono state riprese immagini successivamente in luce verde, rossa, verde e blu. La telecamera con teleobiettivo aveva un solo filtro giallo per eliminare ogni eventuale nebulosità che si presentasse nell'atmosfera.

Mariner 6 fu lanciato senza alcun incidente il 24 febbraio dello scorso anno e *Mariner 7*, 31 giorni piú tardi. *Mariner 7* avrebbe raggiunto Marte soltanto cinque giorni dopo *Mariner 6* poiché la sua rotta era piú corta all'incirca del venti per cento: 316 milioni di chilometri contro 390 milioni.

Cinquanta ore prima del momento di massima vicinanza al pianeta, il CC&S, installato a bordo del *Mariner 6*, ha rivolto le telecamere verso Marte e ha messo in azione un sensore che ha orientato la telecamera con teleobiettivo verso il pianeta. Due ore piú tardi – e da quel momento per altre venti ore – le telecamere hanno fotografato e registrato su nastro analogico 33 immagini del pianeta a intervalli di 37 minuti. Durante questo tempo Marte ha fatto cinque sesti di una rotazione e la distanza tra sonda e pianeta è variata tra 1 241 000 e 725 000 chilometri. Dopo che la registrazione è stata ritrasmessa a Goldstone, il nastro è stato cancellato ed è stata ripresa una seconda serie di immagini a una distanza variabile tra 561 000 e 175 000 chilometri. La registrazione è stata quindi ritrasmessa a Terra e il nastro nuovamente cancellato. Le migliori immagini della seconda serie mostrano strutture di diametro inferiore ai 25 chilometri, con una risoluzione quindi circa sei volte migliore di ogni fotografia ripresa da qualunque telescopio terrestre.

Da queste 50 immagini, le prime nel loro genere, è apparso subito chiaro che Marte assomigliava alle fotografie riprese da Terra tanto quanto queste ultime ai disegni ricavati dall'osservazione visuale. Salvo poche notevoli eccezioni gli osservatori avevano disegnato carte di Marte con netti limiti tra regioni luminose e regioni scure e, in alcuni casi famosi, essi avevano incluso una complessa rete di canali, spesso in forma di linee parallele frequenti e ravvicinate. Tali rappresentazioni erano indubbiamente influenzate dalla tendenza dell'occhio e del cervello umani di includere strutture discontinue ravvicinate in rappresentazioni continue.

Il tranquillo viaggio delle due sonde ha avuto momenti critici poche ore prima che *Mariner 6* iniziasse la serie di riprese ravvicinate in programma. In quel delicato momento in cui tutta l'attenzione era rivolta su *Mariner 6*, *Mariner 7* è rimasto muto con intermittenza per sette ore. Si pensò che dovevа essere stato colpito da un micrometeorite. Quando furono ristabilite le comunicazioni, dall'analisi dei dati relativi alla rotta, si stabilí che il veicolo aveva subito un rallentamento sensibile ed era quindi leggermente fuori rotta. Ma il peggio era che la piattaforma della telecamera era ruotata di diversi gradi e non vi era alcuna possibilità di sapere il suo esatto orientamento se non conoscendo di quanto si era mossa dalla sua precedente posizione.

La situazione fu salvata dal puntamento della telecamera a grandangolo verso Marte che ristabilí la direzione di puntamento dell'intera piattaforma. Lo svolgimento dell'operazione ha fatto aumentare la nostra fiducia nelle capa-

Guardando queste due immagini con uno stereoscopio è possibile avere una visione stereoscopica di Marte. Si tratta di due fotografie a grande distanza riprese da *Mariner 7*, la n. 70 (*qui sopra*) e la n. 69 (*a destra*). La n. 69 è stata ripresa da una distanza di 535 131 chilometri; la n. 70, è stata ripresa 48 minuti piú tardi da una distanza di 514 810 chilometri. La longitudine media delle zone al centro delle immagini è 273° est. Nettamente si staglia l'intera regione attorno a Solis Lacus. Le strutture principali (visibili in entrambe le foto) sono Tithonius Lacus, il «canale» Agathodaemon, Juventae Fons e Aurorae Sinus.

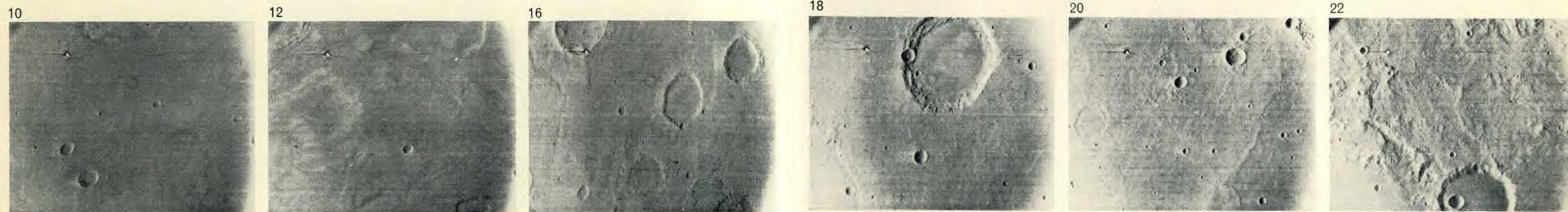
cità operative del sistema che è stato in grado di condurre a buon fine anche questa manovra imprevista. *Mariner 7* ha ripreso tre serie di immagini fotografiche a grande distanza (93 in tutto) in tre periodi di venti ore l'uno, mentre la sua distanza da Marte diminuiva da 1 716 000 a 127 000 chilometri. Nel riprogrammare il sistema CC&S allo scopo di ottenere un'estesa copertura della calotta polare meridionale, si è dovuto tener conto della nuova rotta seguita dalla sonda. L'avvenimento ignoto che aveva provocato il rallentamento della sonda e la modificazione della sua rotta aveva spostato il punto di massimo avvicinamento al pianeta di circa 130 chilometri a sud-est rispetto al previsto. La causa piú probabile di quanto avvenuto sembra oggi essere non tanto l'impatto di un meteorite quanto l'esplosione di una batteria che avrebbe liberato abbastanza gas da mutare la traiettoria del veicolo.

Le fotografie a grande distanza riprese da *Mariner 7* sono state assai utili per l'interpretazione di molte strutture di grandi dimensioni della superficie di Marte. Ad esempio è risultato chiaro che molte delle macchie delle zone meno luminose del pianeta (mari) che appaiono nelle fotografie riprese da Terra, sono associate con un gran numero di crateri i cui diametri raggiungono anche diverse centinaia di chilometri. È apparso inoltre che i limiti dei mari sono graduali in alcune zone e netti in altre; qua e là lunghe impronte scure si estendono nelle zone che da Terra appaiono luminose.

Tali strutture appaiono già anche in diversi disegni e carte ricavate da Terra tramite il telescopio.

È stata compiuta una diligente ricerca di strutture che potessero corrispondere ai canali, a quelle strutture cioè scure, spesso diffuse, piú o meno lineari e generalmente a debole contrasto, alcune delle quali possono essere riconosciute anche in fotografie riprese da Terra. Ne sono un esempio Agathodaemon e Cerberus indicati nelle carte di pag. 26 e 27. Entrambi possono essere riconosciuti nelle fotografie a grande distanza. Altri canali sono invece costituiti da allineamenti di macchie scure di dimensioni e contrasto differenti. Sembra probabile che molti canali non siano altro che il fortuito allineamento di macchie scure distribuite in modo casuale. Forse la struttura piú interessante che compare nel gruppo delle fotografie a grande distanza è la zona luminosa nota col nome di Nix Olympica che è risultata essere un cratere del diametro di oltre 500 chilometri, cioè molto piú grande di qualunque altra nota sulla Luna (*si vedano le illustrazioni in fondo alle pag. 24 e 25*).

Nel considerare le immagini riprese da *Mariner 4* si riteneva che qualcosa o nella telecamera o nell'atmosfera ne avesse ridotto la chiarezza. Qualunque cosa fosse, il fenomeno non si è ripetuto per *Mariner 6* e *7*. Le nuove immagini documentano diversi esempi di sottili strati atmosferici diffondenti che appaiono come stratificazioni parallele al disco del pianeta (*si veda la fig. a pag. 25 in basso, sulla destra*). L'intensità della diffusione atmosferica sem-

In questo mosaico di fotografie a distanza ravvicinata riprese da *Mariner 6* appare una zona di Marte assai ricca di crateri. La distanza media della sonda dalla superficie del pianeta varia da 4920 a 3522 chilometri. Come si può vedere nella carta riportata alle pag. 26 e 27, le immagini 9, 11 e 13 comprendono Margaritifer Sinus, l'« oasi » Oxia Palus e Meridiani Sinus. Le immagini 17, 19, 21 e 23 sono state riprese sopra Sabaeus Sinus e Deucalionis Regio. In molte fotografie riprese da Terra, cosí come in molte di quelle dei Mariner riprese a grande distanza, tali strutture sono caratterizzate da tipiche forme scure. Tuttavia nelle immagini riprese a distanza inferiore il contrasto a grande scala è stato eliminato automaticamente dalle apparecchiature di trattamento delle immagini, mentre è stato esasperato il contrasto a piccola scala, che risulta moltiplicato per tre. Il terminatore è visibile sulla destra dell'immagine 23. I rettangoli neri che appaiono in diverse foto identificano la posizione delle sei immagini ad alta risoluzione riportate qui sotto.

In queste immagini a distanza ravvicinata, la superficie di Marte assomiglia molto a quella lunare; si tratta delle fotografie con teleobiettivo di *Mariner 6* n. 10, 12, 16, 18, 20 e 22 riprese insieme con le immagini grandangolari 11, 13, 19, 21 e 23 del mosaico sopra. Sia quelle che le immagini grandangolari sono state suddivise nello stesso numero di linee (704) e nello stesso numero di elementi per linea (945), ciascuno dei quali contiene lo stesso numero di livelli di luminosità (256). Il teleobiettivo usato per queste foto ha una lunghezza focale dieci volte maggiore di quella utilizzata per le riprese grandangolari e dunque ingrandisce le strutture superficiali di circa 10 volte, rivelando strutture di diametro anche inferiore a 300 m.

bra variare a seconda della posizione e dell'ora del giorno. Il livello diffondente principale ha uno spessore di 10 chilometri e si trova a una quota compresa tra i 15 e i 25 chilometri al di sopra della superficie del pianeta.

Una curiosa caratteristica delle fotografie di Marte riprese da Terra è che quelle di esse scattate con un filtro blu mostrano una notevole diminuzione del contrasto tra le strutture chiare e quelle scure. In rare occasioni tuttavia il contrasto risulta notevolmente esaltato; si è supposto che tale fenomeno, chiamato *blue-clearing*, dipendesse da uno strato diffondente atmosferico della luce blu. Le nuove fotografie riprese dai *Mariner* con filtro blu non differiscono, sotto questo aspetto, da quelle riprese con filtri rosso e verde: la superficie del pianeta è chiaramente visibile con tutti i filtri. Tuttavia fotografie con filtro blu riprese nello stesso momento da Terra mostrano una tipica diminuzione di chiarezza; dunque i *Mariner* non stavano volando verso Marte in uno dei rari periodi di *blue-clearing*. Rimane cosí inspiegato il particolarissimo effetto che si verifica quando Marte viene fotografato da Terra.

Un accurato confronto delle immagini riprese a cinque giorni di distanza dai due *Mariner*, in particolare di quelle in cui l'intero disco del pianeta è visibile, sembra fornire una chiara testimonianza di un'attività atmosferica. Nella regione polare settentrionale, alcune zone che appaiono luminose nelle riprese di *Mariner 6* lo sembrano assai meno in quelle di *Mariner 7*. Poiché le aree riprese sono le stesse, è possibile che il cambiamento di luminosità osservato dipenda dall'entità del gelo superficiale o forse da mutamenti in un'eventuale copertura nuvolosa persistente associata con particolari strutture superficiali. Piú a sud e precisamente in prossimità di Tharsis, Candor, Tractus Albus e Nix Olympica, si possono localizzare aree che divengono luminose al mattino e che raggiungono un massimo di luminosità qualche ora piú tardi (*si vedano le illustrazioni alle pagine 24 e 25*).

Uno degli obbiettivi principali della missione compiuta dai *Mariner* lo scorso anno era quello di fotografare da distanza ravvicinata i principali tipi di strutture superficiali visibili da Terra. *Mariner 4* ci aveva dimostrato che una piccola porzione della superficie, corrispondente a circa l'uno per cento dell'area totale del pianeta, era assai ricca di crateri. Le 58 fotografie da distanza ravvicinata riprese da *Mariner 6* e *7* coprono un'area corrispondente a circa il 10 % della superficie totale del pianeta e ne mostrano un aspetto differente. Oltre a zone ricche di crateri esse ne mostrano altre che sono completamente prive sia di crateri che di altre strutture superficiali e altre ancora a morfologia caotica.

I crateri di Marte sono di due tipi fondamentali: di grandi dimensioni, a fondo piatto ed erosi oppure piccoli, tondeggianti e privi di tracce di fenomeni erosivi. Questi ultimi sembrano crateri d'impatto come alcuni di quelli visibili sulla Luna. Nelle fotografie di Marte tuttavia non è stato visto nulla di simile ai piú grandi e antichi crateri lunari (per esempio il cratere Platone) che sembra siano stati invasi da magma in seguito solidificatosi; cosí come non si è osservato nulla di simile ai raggi che si dipartono dai grandi crateri d'impatto lunari. Anche se è possibile distinguere accumuli di detriti attorno ad alcuni crateri marziani, il fenomeno appare assai meno esteso che non sulla Luna. Le strutture assenti sulla superficie di Marte sono proprio quelle piú facilmente attaccabili dall'erosione o ricopribili; tale osservazione concorda con quella che i crateri mar-

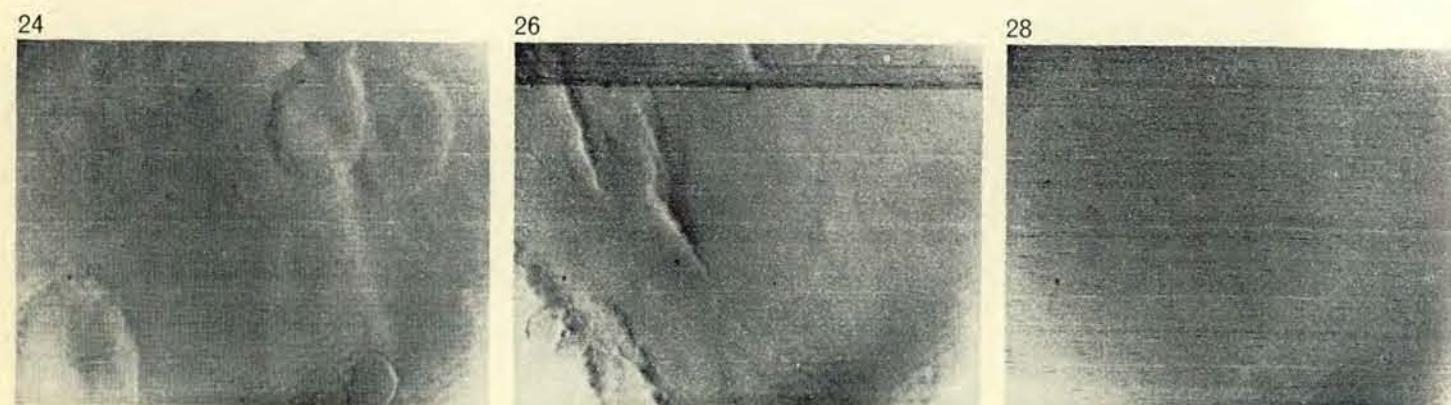
La calotta polare meridionale appare nel mosaico (*in alto*) delle fotografie grandangolari 11, 13, 15, 17 e 19 riprese da *Mariner 7*. Immediatamente qui sopra è riportata la fascia centrale dello stesso mosaico. Mentre però nel mosaico in alto è stata mantenuta la riduzione di contrasto a grande scala introdotta nel trattamento dell'immagine prima dell'invio a Terra, nella fascia immediatamente qui sopra, mediante ulteriore trattamento delle immagini, il contrasto è stato reintrodotto. Appare cosí evidente il notevole contrasto tra la superficie nuda di Marte che appare nella parte sinistra della sequenza e la superficie della calotta polare stessa. La posizione del terminatore tuttavia appare piú evidente nel mosaico in alto (*estrema destra*). Il polo sud marziano si trova presso la serie di dorsali che compaiono nell'angolo in basso a destra del fotogramma 17. Qui sotto sono riportate le sei foto con teleobiettivo riprese contemporaneamente alle foto grandangolari utilizzate per il mosaico in alto.

Queste immagini ravvicinate della superficie di Marte nella zona della calotta polare meridionale mostrano un curioso modo di distribuirsi delle nevi di Marte. Si tratta delle immagini 10, 12, 14, 16, 18 e 20 che sono state riprese contemporaneamente alle grandangolari del mosaico sopra. Il fotogramma 10 mostra la superficie del pianeta fuori della calotta polare; il 12 riprende proprio il margine della calotta. Anche se non sembra cosí il Sole si trova sulla sinistra. I margini di destra dei crateri sembrano in ombra poiché la neve vi è stata rimossa dal Sole. Tuttavia sul picco al centro del maggiore dei quattro grandi crateri sembra vi sia neve: pare quindi che la neve tenda a spostarsi dalle regioni meno elevate a quelle piú elevate.

③

ziani sono generalmente meno profondi e scoscesi di quelli lunari. La sottile atmosfera del pianeta indubitabilmente contribuisce ad accentuare le differenze.

In poche immagini sono visibili dorsali basse e irregolari, simili a quelle che compaiono nei mari lunari nonché alcune strutture lineari assai poco marcate. Tuttavia non è stato possibile identificare con sicurezza nulla che rassomigli ai *rills*, le valli lineari o sinuose a fondo piatto e analoghe alle fosse tettoniche terrestri, che sono stati osservati sulla Luna. Sorprendentemente, inoltre, le fotografie riprese da *Mariner 6* e *7* non sembrano includere nemmeno un esempio delle strutture lineari simili a fratture che sicuramente intersecano uno dei crateri ripresi da *Mariner 4*: fotografando un'area dieci volte maggiore c'era da aspettarsi non uno ma piú esempi di tale struttura. Da ultimo sarà bene osservare che non è stata registrata nemmeno una struttura tettonica analoga a quelle che sulla Terra sono di norma associate con l'orogenesi o con la formazione di archi insulari o con deformazioni dovute a compressione.

Le immagini 25, 27 e 29 riprese da *Mariner 7* hanno mostrato una zona priva di strutture superficiali. È la zona che attraversa Noachis, Hellespontus e penetra nel « deserto » Hellas che talvolta appare piú brillante della stessa calotta polare meridionale. I crateri improvvisamente cessano sulla destra del fotogramma 27; poi, per circa 1600 chilometri, cioè per tutta la estensione di Hellas, non ve ne sono piú. La loro assenza è un ulteriore mistero poiché è altamente improbabile che tale zona abbia potuto per milioni d'anni rimanere priva di crateri. Sotto, su questa e sulla pagina a fronte, ancora piú in dettaglio, la liscia superficie di Hellas (foto 28 e 30).

In queste immagini ravvicinate ciacuna delle quali ritrae la zona compresa nel rettangolo nero del fotomosaico sopra è ben visibile fra l'altro (fotogrammi 28 e 30) la superficie eccezionalmente liscia di Hellas. Qualunque sia il processo erosivo in grado di giustificare la superficie liscia di Hellas, esso sembra esser stato attivo anche nelle zone ritratte dalle due immagini 24 e 26. Queste fotografie sono state riprese mentre *Mariner 7* si trovava alla sua distanza minima dalla superficie del pianeta (3620 chilometri); per questo motivo nelle immagini qui riportate è stato ottenuto il livello di risoluzione piú elevato possibile dell'intera missione di *Mariner 7*.

In un certo senso, i primi studiosi di Marte che definirono l'Hellas un deserto, avevano ragione. L'intera superficie occupata da tale struttura, che ha diametro di circa 1600 chilometri, è completamente priva di crateri o di qualunque altra forma topografica di dimensioni superiori al minimo di 300 metri registrabile dalle nostre telecamere. Sulla Luna non esiste una zona di simili dimensioni altrettanto piatta. Corrispondenti terrestri potrebbero essere i grandi deserti o le grandi pianure.

Le fotografie a grandangolo 25, 27, 29 e 31 riprese da *Mariner 7*, e le corrispondenti con teleobiettivo che hanno numerazione pari, iniziano nella zona nota come Hellespontus e passano attraverso Hellas (*si vedano le illustrazioni sulla pagina a fronte e su questa*). Hellespontus è assai ricco di crateri; vi è poi una zona di transizione a dorsali che ha una larghezza variabile tra i 130 e i 350 chilometri anch'essa ricca di crateri e che sembra declinare dolcemente verso Hellas. Poi, nello spazio di 200 chilometri, i crateri scompaiono del tutto.

Sembra necessario concludere che una volta dovevano esserci i crateri anche all'interno di quest'area e che successivamente essi sono stati obliterati o da fenomeni erosivi o di riempimento tuttora in atto. Ma quale tipo di fenomeno può avere prodotto tale effetto? Può trattarsi di un fenomeno connesso o con la natura del materiale o con agenti locali. Difficilmente si possono immaginare venti cosí localizzati da produrre una copertura di polveri su Hellas (e forse su poche altre località non comprese nelle nostre fotografie) lasciando altrove interi crateri pressoché intatti.

È stata formulata un'ipotesi che, pur essendo, per il momento, priva di qualunque conferma potrebbe dare una risposta a tali interrogativi: l'ipotesi chiamata del *pink-popcorn*. Forse il violentissimo impatto di un asteroide che presumibilmente ha generato l'intera zona di Hellas ha liberato abbastanza calore da scaldare e fondere le rocce circostanti liberando grandi quantità di gas e componenti volatili, creando contemporaneamente grandi quantità di ceneri o materiali pomicei sotto forma di particelle delle dimensioni di un popcorn, abbastanza leggere per essere distribuite dai venti marziani, ma troppo pesanti per essere trasportate al di fuori di quest'area. Niente vieta inoltre che tali ipotetiche particelle siano state caratterizzate da colore rosa dato che tutta Hellas ha spesso un colore nettamente rosato.

Il terzo tipo di morfologia superficiale che abbiamo definito caotica, è stata individuata nella parte centrale della sequenza di fotografie riprese da *Mariner 6* (*si vedano le illustrazioni sulla prossima pagina*). I geologi hanno usato il termine di caotico per descrivere una superficie costituita da dorsali e valli brevi e disordinate. Dalle fotografie si può stimare che tale morfologia interessi un'area di circa un milione di chilometri quadrati, ossia circa la metà dell'area compresa in una sequenza di foto grandangolo che corre parallela all'equatore marziano e a sud di esso. Quest'area, non solo è attraversata da dorsali in tutte le direzioni, ma è anche praticamente priva di crateri.

Le zone a morfologia caotica, inoltre, sono differenti per disposizione e forma dei singoli elementi da quelle regioni lunari che, per gli accumuli di prodotti espulsi dai crateri hanno un aspetto simile. Osservando tali zone marziane si ha l'impressione che gran parte di esse siano sprofondate per la scomparsa di qualche sostanza che ne permeava il sottosuolo, come potrebbe essere uno spesso strato di permafrost. Un'altra possibilità è che nel sottosuolo si sia improvvisamente ritirato del magma o che si sia verificato un qualunque altro fenomeno subsuperficiale associato con il vulcanismo. Quest'ultima ipotesi sembra però essere esclusa dall'assenza in superficie di caratteristiche strutture vulcaniche.

Non c'era da attendersi che le fotografie della calotta polare meridionale marziana ponessero fine alle discussioni tra coloro che ritengono che tali calotte siano costituite da neve e ghiacci normali e coloro che al contrario le ritengono costituite da anidride carbonica solida (ghiaccio secco). Coloro che hanno condotto e seguito la ripresa delle fotografie ritengono che tutte le conclusioni che si possono trarre dalle immagini favoriscano l'ipotesi del ghiaccio secco.

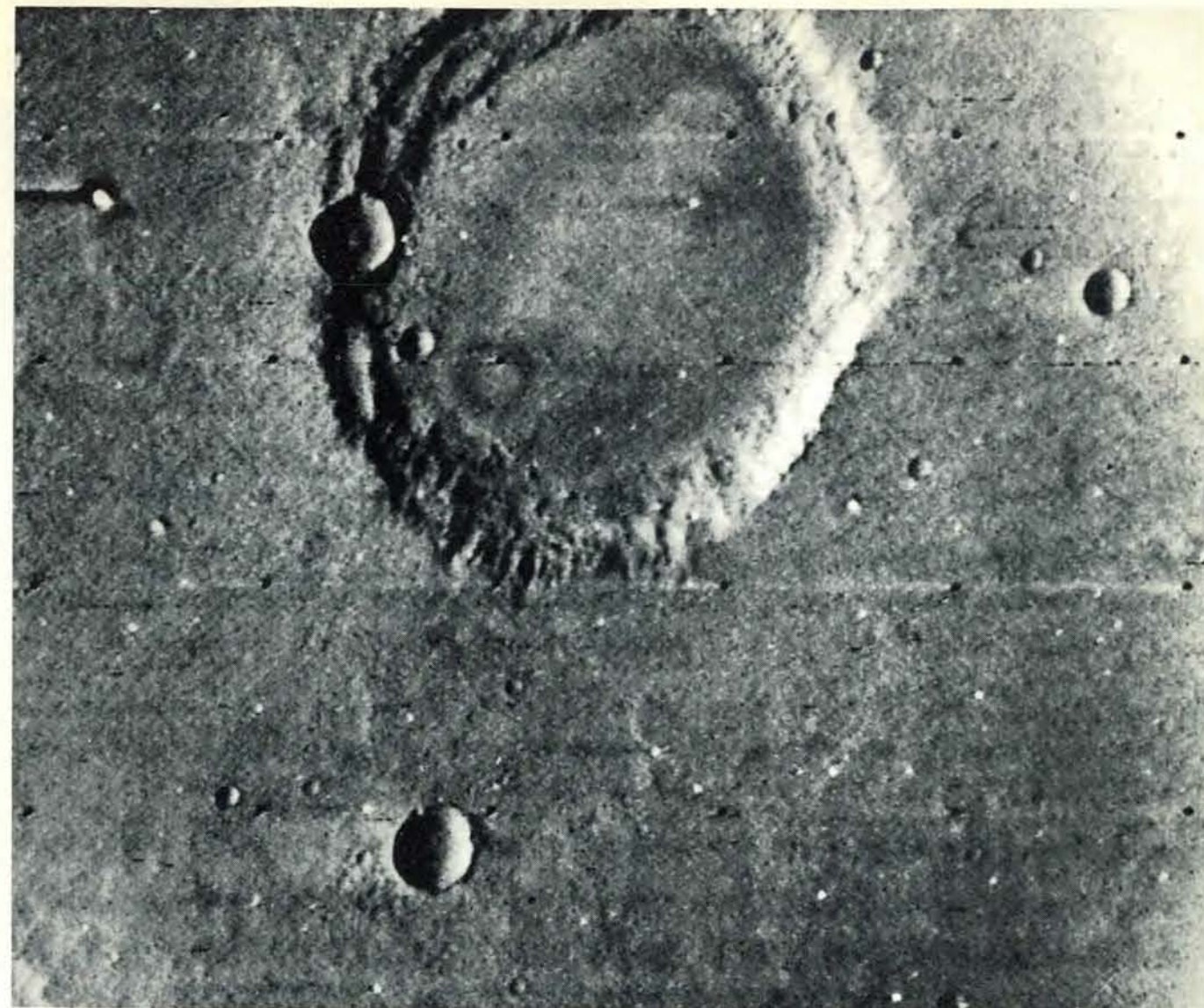
La brillantezza della calotta polare cosí come è stata osservata potrebbe essere dovuta alla copertura di un qualunque materiale polverulento bianco. D'altro canto certi tipi di rilievo osservati nelle riprese della calotta polare e assenti fuori di essa suggeriscono che la neve (se di neve si tratta) deve essere stata ammucchiata in cumuli di diversi metri di altezza. Nello stesso tempo la morfologia del margine della calotta polare indica una maggiore influenza dell'evaporazione dovuta al calore solare che non del vento.

Supponendo che l'evaporazione sia dovuta principalmente al riscaldamento di mezzogiorno si può stimare che la

Altrettanto notevole e inaspettata quanto la superficie liscia di Hellas, è la morfologia caotica messa in evidenza dalle fotografie di *Mariner 6* che coprono una striscia lunga circa 2000 chilometri poco a sud dell'equatore marziano, nella regione di Aurorae Sinus. Nei fotogrammi 14 *(in alto)* e 8 *(in basso)* si possono osservare dettagli delle zone a morfologia caotica. La zona compresa nella foto 14 è lunga circa 235 km e larga 100; quella ripresa in 8 è lunga 126 e larga 97 km. Le dorsali disordinate che appaiono in queste immagini non hanno un corrispondente né sulla Luna né sulla Terra.

evaporazione netta giornaliera sia di circa un grammo per centimetro quadrato per l'anidride carbonica solida e di solo un decimo di tale valore per il ghiaccio normale. Dato che l'evaporazione di tutta la calotta a una certa latitudine richiede molti giorni, si può moltiplicare l'evaporazione netta giornaliera per il numero di giorni di primavera e d'estate (almeno 100) ottenendo cosí una stima dello spessore minimo totale della calotta. Il risultato è di decine di grammi per centimetro quadrato nel caso si tratti di anidride carbonica solida e, invece, di alcuni grammi per centimetro quadrato se si tratta di ghiaccio normale.

Ci si potrebbe ora chiedere se l'atmosfera marziana, composta probabilmente soprattutto da anidride carbonica con uno scarso contenuto di vapor d'acqua, è in grado di trasportare e di concentrare l'una o l'altra di queste sostanze dall'uno all'altro polo nell'intervallo tra l'inverno e l'estate. Si può concludere che se la piccola quantità d'acqua indicata dai calcoli sull'evaporazione deporrebbe a sfavore della possibilità che l'atmosfera marziana ne contenga anche in piccola quantità, molto meno ancora ne dovrebbe poter trasportare da un polo all'altro. Tale trasporto non è invece necessario per l'anidride carbonica poiché l'atmosfera marziana è ricca di tale gas: l'atmosfera semplicemente dovrebbe cederne un poco all'uno o all'altro polo a seconda della stagione. Sembrerebbe logico dunque concludere che le nevi di Marte siano almeno in larga misura, se non interamente, costituite da anidride carbonica solida, cioè da ghiaccio secco.

Un problema sorto dall'analisi delle fotografie della calotta polare è che gran parte dei crateri hanno fondo scuro e versanti rilucenti. Questo è esattamente quanto ci si sarebbe aspettato sulla Terra dove la neve tende a cadere e ad accumularsi alle altitudini piú elevate. Su Marte tuttavia le precipitazioni dovrebbero verificarsi soprattutto alle quote inferiori ove la pressione atmosferica è piú elevata. Questa considerazione è servita di stimolo nella ricerca di un meccanismo che potesse trasportare la neve dalle quote inferiori a quelle piú elevate. Benché anche il vento sia un possibile agente di trasporto dell'anidride carbonica solida, è piú probabile che essa evapori dal fondo dei crateri ove si depone, per ridepositarsi piú in alto dove la temperatura è inferiore. Un'altra possibilità è che il fondo di alcuni crateri appaia scuro perché ricoperto di un sottile strato di ghiaccio come accade sulla Terra per un qualunque bacino lacustre in

In questa immagine ripresa da *Mariner 6* (fotogramma 18) Marte assomiglia molto alla Luna. Per apprezzare i dettagli registrati dalla telecamera è bene paragonare questa immagine con quella riportata a pag. 31 in basso. La morfologia a fondo piatto del grande cratere è tipica di quelli che si ritiene siano i piú antichi; i piú giovani invece hanno un fondo arrotondato. Questa immagine mostra crateri grandi un terzo dei piú piccoli crateri visibili sulla Luna mediante i telescopi terrestri.

tali condizioni, che appare nettamente scuro quando viene osservato dall'alto.

Come tutti si attendevano le fotografie riprese da *Mariner 6* e *7* non hanno fornito alcuna indicazione diretta dell'esistenza di vita su Marte. Se per caso ci fosse vita su Marte è molto probabile che sarebbe unicellulare e quindi del tutto invisibile con i mezzi utilizzati. Ci si potrebbe ancora chiedere se allo stato attuale delle conoscenze si ritiene ancora possibile l'esistenza di vita sul pianeta rosso. Dall'esame delle fotografie di Marte si potrebbe concludere che l'evidente scarsità di acqua è il fattore piú importante che si oppone all'esistenza di vita nella forma che noi conosciamo. L'abbondanza di crateri e l'assenza di strutture tettoniche analoghe a quelle terrestri sembra indicare che Marte non ha mai avuto oceani delle dimensioni di quelli terrestri per un tempo lunghissimo e che è possibile che non ne abbia mai avuti. Ancora non è possibile dire quanto grande debba essere e per quanto tempo debba sussistere un oceano perché in esso la vita possa iniziare. Se Marte deve essere l'occasione per saggiare le nostre cognizioni sull'origine della vita dobbiamo evitare di usare le stesse nozioni per concludere, prima ancor di cominciare, che Marte è privo di forme di vita.

Non si è ancora accennato ad alcuni dei piú importanti risultati che ci si attendeva dall'esame delle fotografie di Marte: quelli che derivano dalla misura quantitativa della luminosità e dalle relazioni geometriche delle strutture riprese nelle diverse immagini. Il processo di analisi delle immagini, che eventualmente fornirà tali dati, è ancora in fase sperimentale e occorrerà molto tempo e grande cura perché possa essere utilizzato con successo. I risultati che eventualmente ne potranno sortire saranno una migliore conoscenza della forma e delle dimensioni del pianeta, una carta di Marte vincolata a punti topografici di riferimento fissi, una piú precisa definizione della direzione dell'asse di rotazione, misure quantitative delle pendenze topografiche, una migliore conoscenza dei colori e della riflettività dei materiali superficiali, dei fenomeni di diffusione e di assorbimento della luce nell'atmosfera marziana. Le fotografie di *Mariner 6* e *7* saranno per lungo tempo una pietra miliare nella conoscenza della storia passata e delle condizioni attuali dello spazio che ci circonda.

Lo zolfo

Dal 1900 le richieste dell'agricoltura e dell'industria moderne hanno aumentato di quaranta volte il consumo di questo elemento. Esso proviene da molte fonti e circa l'87% è trasformato in acido solforico

di Christopher J. Pratt

L'economia degli Stati Uniti consuma zolfo per circa 45 chilogrammi annui a persona e ciononostante l'abitante medio non si accorge dell'esistenza di questo elemento. Lo zolfo ha modi invisibili per denunciare la sua presenza: è usato come acido solforico nella fabbricazione di fertilizzanti, tessili, materie plastiche, coloranti e molti altri prodotti; come parte essenziale dei fertilizzanti; come ingrediente in medicina e infine è presente anche come inquinante atmosferico sotto forma di anidride solforosa. La probabilità di venir a contatto con lo zolfo era comunque superiore nei tempi antichi, quando lo zolfo veniva bruciato comunemente per scacciare gli spiriti maligni, tener lontani gli insetti, sbiancare le stoffe e disinfettare le case dopo le malattie. Un viaggiatore del XV secolo, Leo Africanus, si riferisce ai bruciatori di zolfo (in particolare agli alchimisti) come a « persone stupide che si mummificano con lo zolfo e altri terribili puzze ».

Lo zolfo è un elemento a carattere non metallico, con forte tendenza a combinarsi con gli altri elementi. Nella sua forma elementare si trova in depositi sedimentari o vulcanici: non elementare, combinato con uno o più elementi, si trova principalmente nei solfuri e nei solfati. Lo zolfo è di per sé inodoro, contrariamente alla comune esperienza dell'odor pungente dell'anidride solforosa e di quello di uova marce dell'idrogeno solforato, ed è anche insaporo. È un isolante per il calore e l'elettricità, è insolubile nell'acqua e nella maggior parte degli acidi, ma solubile nel solfuro di carbonio e in parecchi altri liquidi organici.

Una misura della sua importanza in economia è data dalla produzione annuale di zolfo in Canada e negli USA che ammonta a 13 milioni di tonnellate. Circa l'87% della produzione va in acido solforico, tradizionale cavallo di battaglia dell'industria chimica; lo si usa più di tutti gli altri intermedi nei vari processi chimici. Quantità apprezzabili di zolfo vengono bruciate per formare anidride solforosa da usare in svariati processi, compreso quello della fabbricazione di acido solforico. A causa della quantità di zolfo consumato e all'ampio spettro dei suoi usi, questo elemento è spesso considerato — assieme al salgemma, al calcare, al carbone e al petrolio — una delle cinque materie prime fondamentali per l'industria chimica.

Proprietà dello zolfo

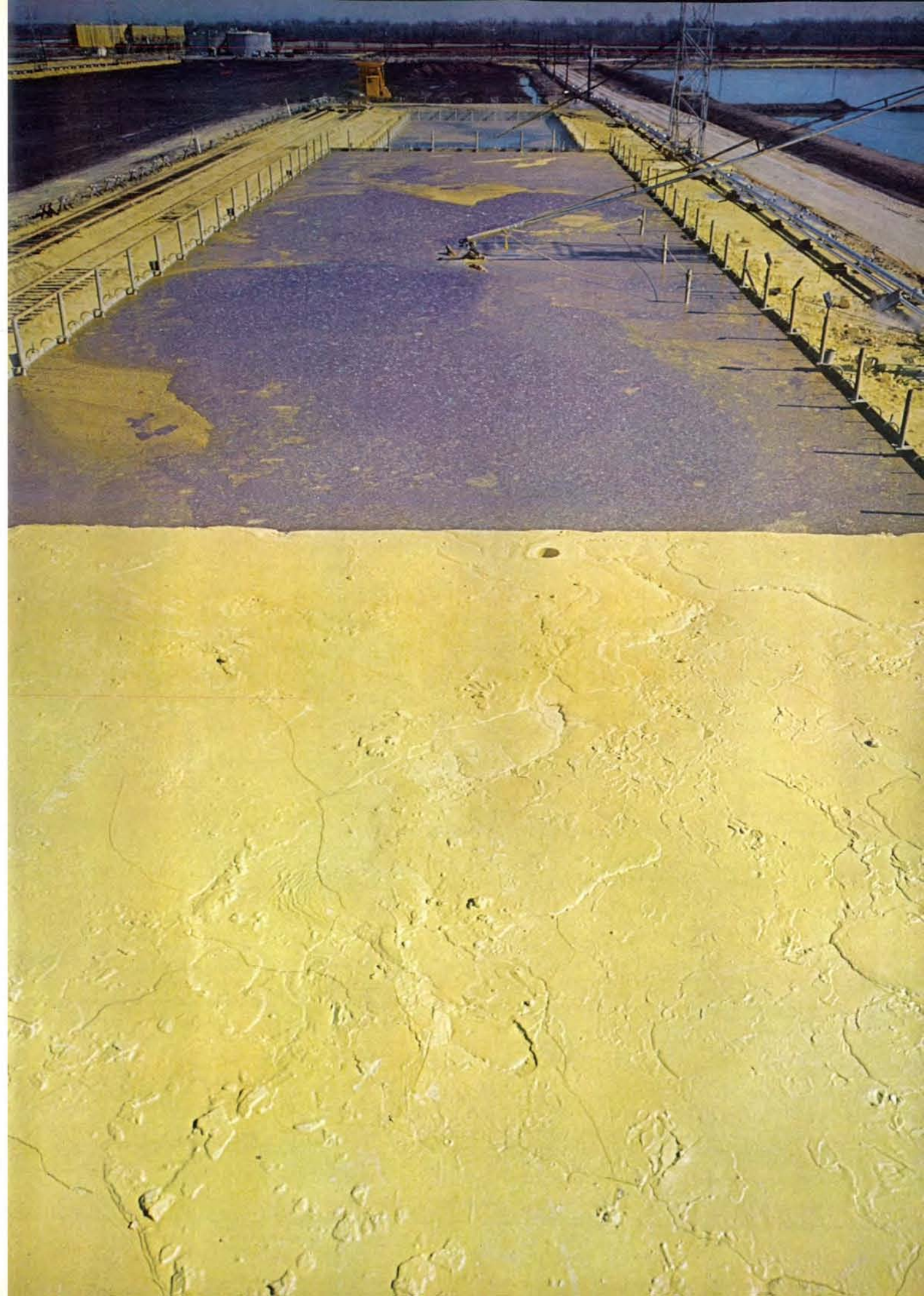
Nella sua forma elementare lo zolfo può assumere una varietà insolita di configurazioni molecolari e di forme cristalline. La struttura molecolare base è un anello distorto a otto atomi: sedici di questi anelli costituiscono la cella elementare della forma cristallina che va sotto il nome di zolfo rombico e sei quella che va sotto il nome di zolfo monoclino, instabile. Si conoscono ben 36 forme cristalline dello zolfo.

Il raffreddamento e il riscaldamento dello zolfo danno luogo a proprietà insolite. Un improvviso raffreddamento dello zolfo fuso produce un materiale gommoso in cui sono stati riconosciuti grappoli di catene atomiche a otto atomi per catena. Dopo il raffreddamento le catene si trasformano lentamente in cristalli rombici stabili. Se lo zolfo rombico viene riscaldato esso si trasforma reversibilmente in zolfo monoclino alla temperatura di 95,4 °C e quest'ultimo fonde a 111,5 °C, formando un liquido giallo paglierino con strutture a otto atomi. Un ulteriore riscaldamento fa diventare il liquido maggiormente fluido, mentre il colore si incupisce.

A circa 160 °C, però, il liquido diventa improvvisamente molto vischioso, come una colla che si rapprende. Questo fenomeno sembra si debba attribuire alla formazione di polimeri a catena lunga. Se nello zolfo sono presenti impurezze organiche in piccola quantità, oppure cloro o iodio, i cambiamenti di viscosità si riducono in quanto queste sostanze inibiscono la formazione di polimeri.

La viscosità persiste sino a 230 °C, mentre a temperature più elevate diminuisce, ma il colore scuro permane sino al punto di ebollizione, cioè sino a 444,6 °C. Un ulteriore riscaldamento del vapore porta a strutture a due e a un atomo. Questi fenomeni sono ottenuti in senso inverso per lento raffreddamento.

L'alta reattività dello zolfo permette l'esistenza di un gran numero di composti naturali e sintetici in cui l'elemento è legato in maniera covalente a uno o più atomi. (La covalenza si riferisce al numero di elettroni condivisi tra atomi ed è una misura della reattività chimica. Mentre nella maggioranza dei composti l'ossigeno mostra covalenza due, lo zolfo ha covalenza due, tre, quattro e sei.) Lo zolfo si combina con quasi tutti gli elementi, fatta eccezione per l'oro, il platino e i gas inerti. Tra i composti più noti e utili vanno segnalati l'idrogeno solforato (H_2S), l'anidride solforosa (SO_2) l'acido solforico (H_2SO_4) e sostanze co-

Un tino di zolfo in costruzione nell'impianto della Texas Gulf Sulphur Company a Newgulf nel Texas. Il materiale violaceo nella parte superiore della fotografia della pagina a fronte è zolfo fuso appena spruzzato nel tino. Il materiale giallo nella parte inferiore è invece zolfo già solidificato. Entrambe le parti della fotografia mostrano la parte superiore del tino, anche se lo zolfo solidificato è alto circa un metro di più dello zolfo fuso. In alto a sinistra sullo sfondo vi è il tino che appare sulla copertina.

L'« esodo volontario » di 1588 lavoratori occupati nel settore e la previsione che anche dopo la riorganizzazione (per la quale è previsto un investimento di 34 miliardi) lo zolfo siciliano verrà a costare un po' piú del doppio dello zolfo di importazione, sono gli ultimi atti del dramma della nostra industria solfifera. La produzione globale nazionale di zolfo minerale continua a scendere (541 466 tonnellate nel 1968, 419 068 nel 1969) mentre, malgrado la sopravvivenza del regime protezionistico della produzione di zolfi nazionali, l'importazione di zolfi greggi continua ad aumentare (120 492 tonnellate nel 1968, 175 777 nel 1969). È un dramma che trova la sua dimensione umana nelle oltre tremila persone che in tale settore sono direttamente impiegate (3259 operai al dicembre 1969). Lo zolfo siciliano (che rappresenta la parte preminente della disponibilità nazionale è concentrato in determinati livelli di una formazione geologica vecchia di 20 milioni di anni per l'appunto chiamata formazione gessoso-solfifera (Miocene superiore). Secondo uno studio di P. Mezzadri del 1964, lo zolfo siciliano sarebbe stato deposto per semifiltrazione di idrocarburi solforati nel corso della loro migrazione dalle rocce madri sottostanti verso livelli superiori; i livelli filtranti (gessi, argille e calcari cavernosi) avrebbero trattenuto lo zolfo, lasciandosi invece attraversare dagli idrocarburi leggeri. La connessione tra deposizione dello zolfo e migrazione degli idrocarburi (che avviene di preferenza lungo linee di frattura) spiega anche il fatto che i giacimenti siciliani tendano a esser decisamente piú frequenti in zone disturbate dal punto di vista tettonico. E, d'altro canto, questa è una delle ragioni fondamentali per cui non è applicabile ai nostri giacimenti il metodo di estrazione Frasch che ha fatto la fortuna dei giacimenti americani.

Le prime coltivazioni minerarie siciliane risalgono al 200 a.C.: allora lo zolfo era usato soprattutto in medicina; poi, con i romani prima e con l'invenzione della polvere da sparo poi, entra nell'uso bellico ed è cavato con lena sempre maggiore. Quando nel 1787 viene scoperto il processo Leblanc per la fabbricazione della soda la Sicilia è pressoché l'unica regione in grado di fornire zolfo in grande quantità; segue poi la produzione di sempre maggiori quantità di acido solforico: l'industria mineraria siciliana fiorisce. Piú che altro però fioriscono i proprietari dei terreni minerari che, se da un lato percepiscono il cosiddetto « estaglio » che arriva fino al 30% della produzione, dall'altro non gestiscono mai direttamente le miniere; ne nasce di conseguenza che i gestori, i quali per contratto devono alla loro cessazione lasciare ai proprietari tutti gli impianti da loro eventualmente costruiti, sfruttano al massimo le miniere con i mezzi piú primitivi. Comunque tutto va a meraviglia sino alla prima grave crisi, determinatasi per l'accumulo di giacenze, quella del 1883. Nel frattempo (1835) viene messo a punto il primo forno per l'arrostimento delle piriti che consente ai produttori di acido solforico di utilizzare pirite e non zolfo come materia prima. Ne potrebbe derivare una grave crisi se non fosse per la scoperta dell'utilità dello zolfo in agricoltura soprattutto per combattere l'oidio della vite, che dà nuovo tono all'estrazione: nel 1893 si raggiunge la vetta delle 375 000 tonnellate. Già da due anni in America Hermann Frasch tenta di applicare il suo metodo di estrazione: nel 1898 tenterà di applicarlo anche in Sicilia, ma sarà un fallimento. Nel 1895-96 nuova crisi e nuovo accordo con industriali stranieri: nasce l'Anglo-Sicilian Sulphur Co. La produzione rifiorisce: nel 1901 si raggiungono le 537 000 tonnellate e l'impiego di 39 000 unità lavorative. Nel 1906 la Anglo-Sicilian si scioglie. Per superare i contrasti e la concorrenza interna fra i singoli produttori nasce il Consorzio obbligatorio per l'industria solfifera siciliana con lo scopo fondamentale di stabilire accordi per la spartizione dei mercati con la Union Sulphur americana, che è già divenuta una temibile concorrente. Nel 1911 si vendono 534 000 tonnellate. Poi la guerra e i minatori al fronte: nel 1917 la produzione scende a 186 500 t. Nel frattempo sono scaduti gli accordi con la Union Sulphur che, per il calo dei noli marittimi, può partire alla conquista dei mercati piú lontani: alla fine degli anni venti la produzione siciliana si stabilizza sulle 240 000 t. Nel 1933 nasce l'Ufficio per la vendita dello zolfo italiano che nel 1940 si trasforma nell'Ente Zolfi Italiani. Intanto, nel 1927, è stata dichiarata la demanialità dei sottosuoli minerari che sottrae (ma troppo tardi) almeno quelli alle gabelle dei proprietari dei terreni. Dopo la seconda guerra mondiale l'industria mineraria siciliana conosce un solo momento felice, il 1951, un anno di gravi difficoltà dell'industria americana. Poi la produzione americana riprende forza, mentre vengono scoperti giacimenti in altre parti del mondo (Messico, Canada, Francia, Polonia) e mentre lo zolfo di recupero della raffinazione degli idrocarburi assume sempre maggiore importanza. Nel periodo gennaio-ottobre 1969 le esportazioni sono di 548 t. di zolfi greggi e di 5720 t. di zolfi raffinati.

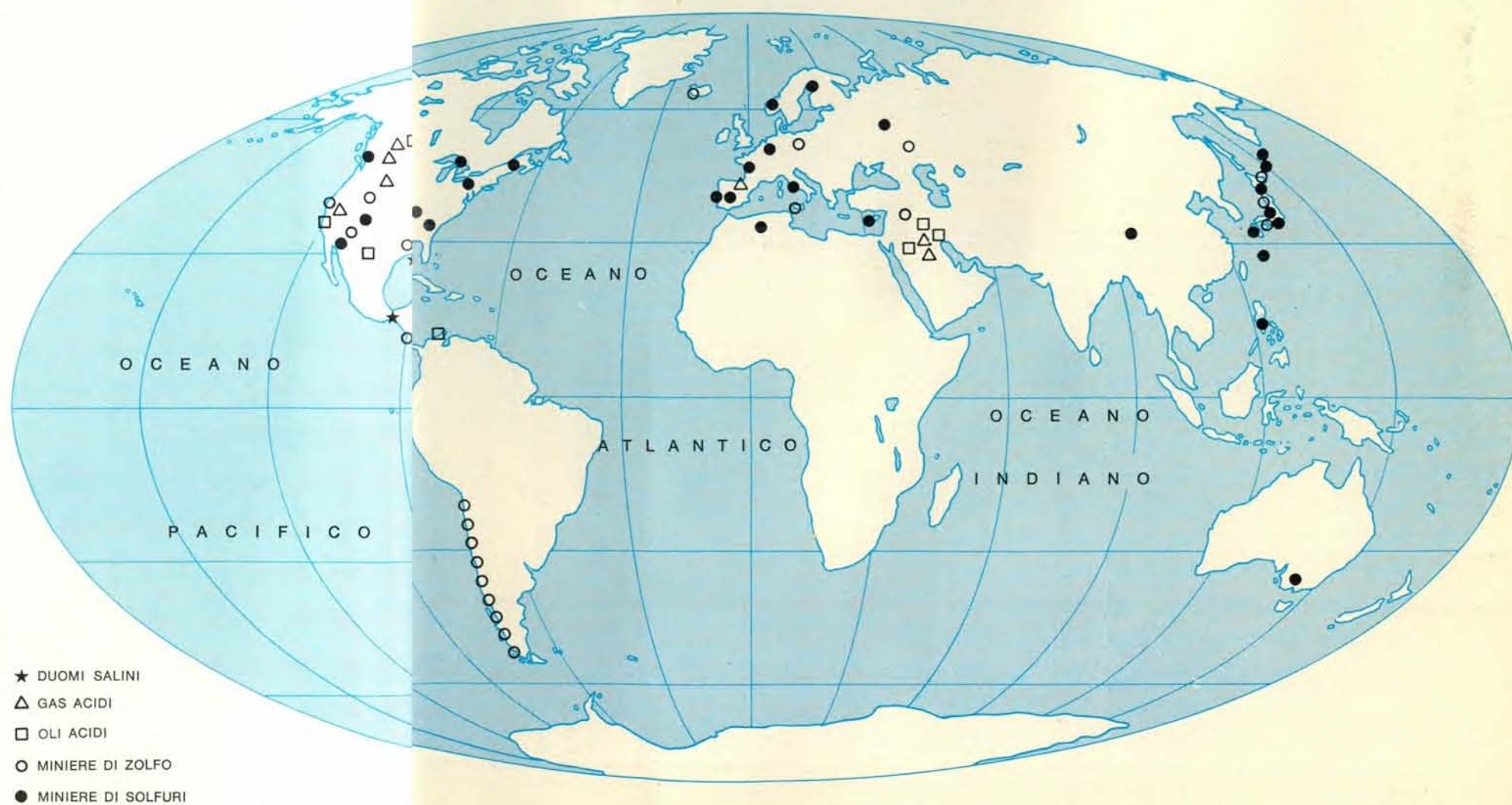
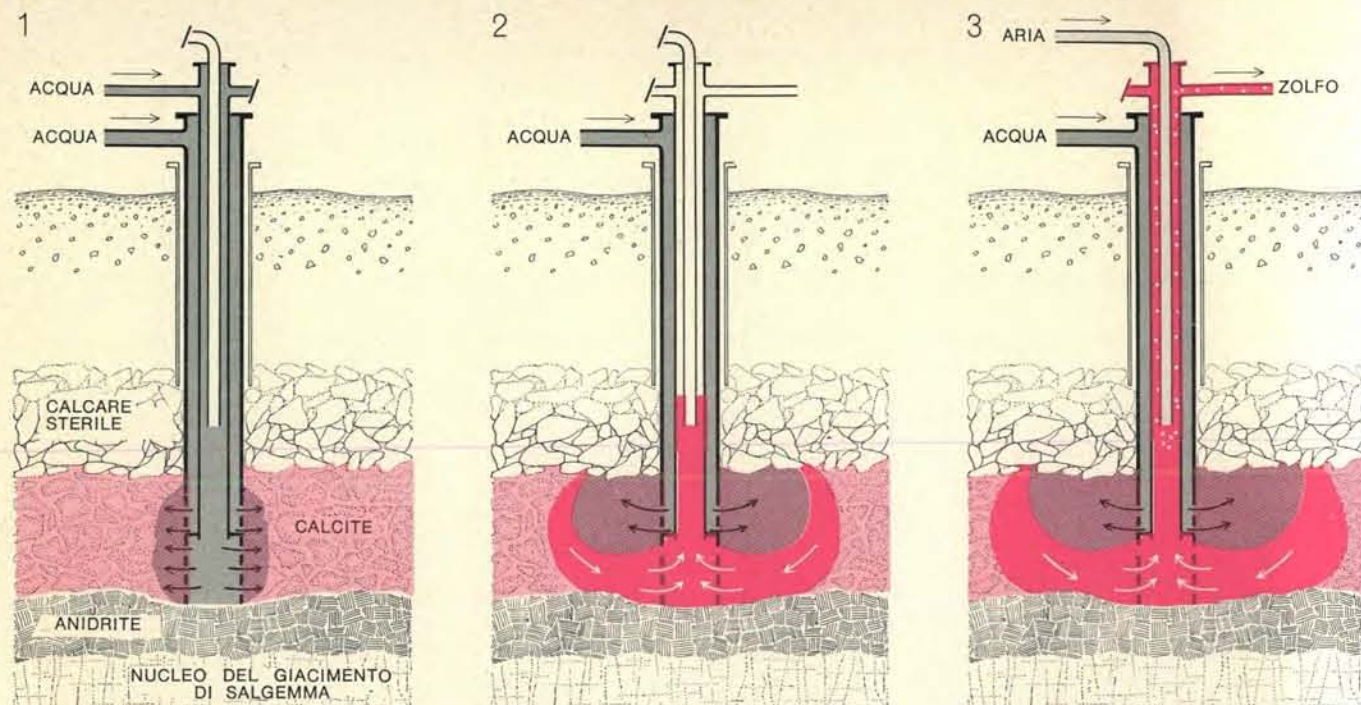
me il sale di Glauber (solfato sodico) e il gesso (solfato di calcio).

È piuttosto difficile classificare i derivati dello zolfo a causa delle sue molteplici forme e della sua larga reattività. La maggior parte, però, può esser suddivisa in composti molecolari e non, in base alle loro proprietà strutturali. L'anidride solforosa è un esempio di derivato molecolare, ed è ottenuta bruciando zolfo in aria. I polisolfuri, le forme polimorfiche dello zolfo e molti composti organici appartengono a questa stessa categoria. I composti non molecolari comprendono i sali ionici solubili in acqua, come certi solfati, e quelli non solubili come per esempio i solfuri di piombo e argento. I composti insolubili in acqua sono cristallini e hanno la struttura teoricamente infinita propria dei reticoli ionici cristallini. Una parte dei composti noti piú insolubili è costituita da questo tipo di disposizioni atomiche e perciò molti solfuri hanno resistito alla lisciviazione naturale occorsa nei secoli.

Sorgenti dello zolfo

Lo zolfo è diffuso sulla Terra in forme elementari e non. È possibile che entrambe siano state originate, all'inizio della storia della Terra, da emissioni di lava e gas contenenti zolfo. Il fenomeno è ancora osservabile: per esempio in certe parti del Giappone e del Cile, l'idrogeno solforato proveniente da sfoghi vulcanici si ossida all'aria e continua a formare depositi locali di zolfo che possono assumere anche valore commerciale. Zolfo nativo si trova anche in prossimità di alcune sorgenti calde.

Le maggiori riserve sono costituite, però, da depositi di sedimentazione. In molti di questi lo zolfo è associato al gesso (solfato di calcio biidrato, $CaSO_4 \cdot 2H_2O$), alla anidrite (solfato di calcio anidro, $CaSO_4$) o alla calcite (carbonato di calcio, $CaCO_3$). Le riserve di tipo sedimentario costituiscono i depositi di zolfo elementare che si trovano in Sicilia, Polonia, Iraq, Texas e nella regione del Golfo del Messico. Lo zolfo combinato dei gas naturali della Francia del sud, del Canada occidentale e di molti paesi del Medio Oriente è anche, probabilmente, di origine sedimentaria, cosí come lo sono i petroli grezzi acidi dei Caraibi e del Medio Oriente e numerosi giacimenti di carbone ad alto tenore di zolfo. La maggior parte dello zolfo presente negli oceani si trova sotto forma di sali disciolti. Lo zolfo non è uno degli elementi piú diffusi sulla crosta terrestre: ha una abbondanza media dello 0,06 %, contro quella dell'8 % dell'alluminio, del 4 % del ferro, del 3,4 % del calcio, dello 0,005 % del rame, dello 0,001 % del boro. (Lo zolfo è stato trovato anche nelle rocce lunari portate sulla Terra dagli astronauti dell'Apollo. L'analisi, effettuata dall'Istituto di geofisica e di fisica planetaria della Università di California a Los Angeles, ha mostrato che il tenore in zolfo, in sei campioni, va da 230 a 650 parti per milione.)

Una delle ragioni della presenza dello zolfo nelle rocce sedimentarie, può esser stata la riduzione del solfato di calcio, come gesso o anidrite, da parte di materiali organici con l'aiuto di opportuni batteri. C'è qualche evidenza che i solfati di sedimentazione, originati forse da reazioni tra l'ossido di calcio di rocce fuse e le emissioni sulfuree vulcaniche, siano stati ridotti dal metano e da batteri anaerobici, come il *Desulfovibrio desulfuricans*, a idrogeno solforato e carbonato di calcio. Secondo questa ipotesi si assume che il materiale organico occorrente e gli altri fattori responsabili provenissero da materiali cellulari in decomposizione, presenti nelle antichissime paludi e foreste. La susseguente ossidazione dell'idrogeno solforato (per mezzo di reagenti chimici naturali, ma anche a opera di batteri come il *Thiobacillus thioparus*) ha portato presumibilmente ai depositi di zolfo elementare. Fin dall'antichità si sapeva che lo zolfo può esser accompagnato da idrocarburi, come lo sono il bitume, il petrolio e i gas naturali; que-

La figura rappresenta le località della Terra dove si trovano le piú importanti scorte di zolfo. La struttura geologica in cui si trovano i giacimenti di zolfo è di solito un deposito di salgemma in cui lo zolfo elementare è presente nelle rocce che si trovano al di sopra del deposito. Questo è il caso, per esempio, dello zolfo che viene estratto a Newgulf nel Texas. Negli altri giacimenti riportati, lo zolfo si trova combinato con altri elementi come, per esempio, carbonio, idrogeno e ferro.

Il processo Frasch è usato per estrarre lo zolfo dalle rocce sovrastanti un giacimento di salgemma. Il processo comporta la sistemazione di tre tubazioni concentriche in una camicia che permette di affondare una conduttura forata nel deposito di zolfo. I diametri che si usano normalmente per le tre tubazioni sono di 8, 4 e 1 pollice. Dell'acqua, riscaldata molto al di sopra del proprio punto di ebollizione, è pompata (*1*) nella camicia tra le tubazioni da 8 e 4 pollici e anche, almeno inizialmente, nella conduttura da 4 pollici. L'acqua fuoriesce dalla conduttura forata nelle rocce e fonde lo zolfo (*2*) che forma una sorta di polla al fondo del pozzo e sale, sotto la pressione dell'acqua, per una certa altezza nella tubazione da 4 pollici. Dell'aria compressa spinta in basso nella tubazione da 1 pollice (*3*) rende più leggero lo zolfo, e lo fa salire in superficie.

sta associazione non meraviglia dato che la tappa fondamentale in ognuno di questi casi è l'ossidazione anaerobica dei materiali organici, assieme alla presenza di metano. Infatti la presenza di zolfo o idrogeno solforato contenente certi rapporti isotopici specifici, indica spesso presenza di petrolio. Si ritiene, ormai, che lo zolfo si sia formato epigeneticamente, cioè dopo consolidamento, e non singeneticamente, cioè durante la formazione del giacimento.

In aggiunta ai grandi giacimenti di zolfo elementare o combinato, esistono depositi grandi di zolfo non elementare sotto forma di minerali sulfurei di ferro, rame, piombo e altri metalli.

Il fatto che questi minerali siano molto poco solubili in acqua, per cui è difficile solubilizzarli e trasportarli in giacimenti di sedimentazione, suggerisce che la loro origine sia vulcanica. I solfuri e in particolare quello di idrogeno, presente nei gas naturali, forniscono considerevoli quantità di zolfo.

Metodi di preparazione

Sino alla fine del XIX secolo la maggior parte del fabbisogno di zolfo proveniva dai giacimenti siciliani, dai quali lo zolfo era estratto con metodi simili a quelli impiegati nelle miniere di carbone. Il minerale veniva purificato riscaldandolo in forni appositi o nei calcaroni; per riscaldamento lo zolfo fondeva e scolava via dalle impurezze del minerale. Questo tipo di processi trova ancora un uso limitato, assieme ad altri più moderni, per ottenere lo zolfo dai suoi minerali: estrazione con solvente organico come tetracloruro di carbonio o toluolo; flottazione a schiuma con uso di agenti che permettano la separazione dello zolfo dal resto del materiale; infine, fusione con vapore o acqua surriscaldata.

Sin dalla scoperta del petrolio si venne a conoscere una enorme sorgente potenziale di zolfo, ma per il suo sfruttamento fu necessario aspettare il superamento di alcune difficoltà di produzione. Alcuni perforatori, in Luisiana, avevano scoperto dei giacimenti di salgemma, attraversati da una vena di zolfo dello spessore di 30 metri, nella parte di rocce sovrastante il giacimento. Le carote prelevate nella perforazione mettevano in evidenza circa 120 metri di argilla e ghiaia, quindi sabbia attraversata da correnti di acqua sulfurea, seguita da un letto di circa 200 metri di calcare riempito di zolfo e, finalmente, uno strato di gesso. Ripetuti sforzi per mettere a giorno il giacimento di zolfo con lavoro in miniera, furono frustrati dal deposito sabbioso e dalle correnti sotterranee di acque sulfuree, assieme a tasche gassose di idrogeno solforato, piuttosto pericolose.

La soluzione a questi problemi venne data nel 1890 da Herman Frasch, un chimico petrolifero. Dato che lo zolfo rombico fonde a temperatura relativamente bassa, egli propose di fonderlo nelle rocce in cui si trova e di pomparlo fuori attraverso tubazioni, così come si fa con il petrolio o l'acqua salata. Il suo suggerimento per fondere lo zolfo consisteva nell'inviare nelle rocce acqua surriscaldata.

Il metodo Frasch, provato per la prima volta nel 1894, si dimostrò di successo: con esso si ottiene oggi la maggior parte dello zolfo nell'America del nord e una buona parte di quello dei giacimenti del Golfo del Messico. Un pozzo Frasch (*si veda l'illustrazione in alto nella pagina a fronte*) si costruisce affondando una custodia protettiva nella parte superiore del letto di anidrite e gesso che si trova sotto lo strato che porta lo zolfo. In questa custodia si introducono tre tubazioni concentriche. Dapprima si immerge nella custodia una conduttura, del diametro di 8 pollici (20 cm), che attraversa le rocce sovrastanti il giacimento e che arriva al fondo del deposito di zolfo. La parte inferiore di questa conduttura è forata, a piccoli fori. Nell'interno di questa conduttura si abbassa un tubo di 4 pollici (10 cm), fin quasi a toccare il fondo. Infine si ha una tubazione di 1 pollice (2,5 cm), la più interna, che arriva circa alla metà del pozzo.

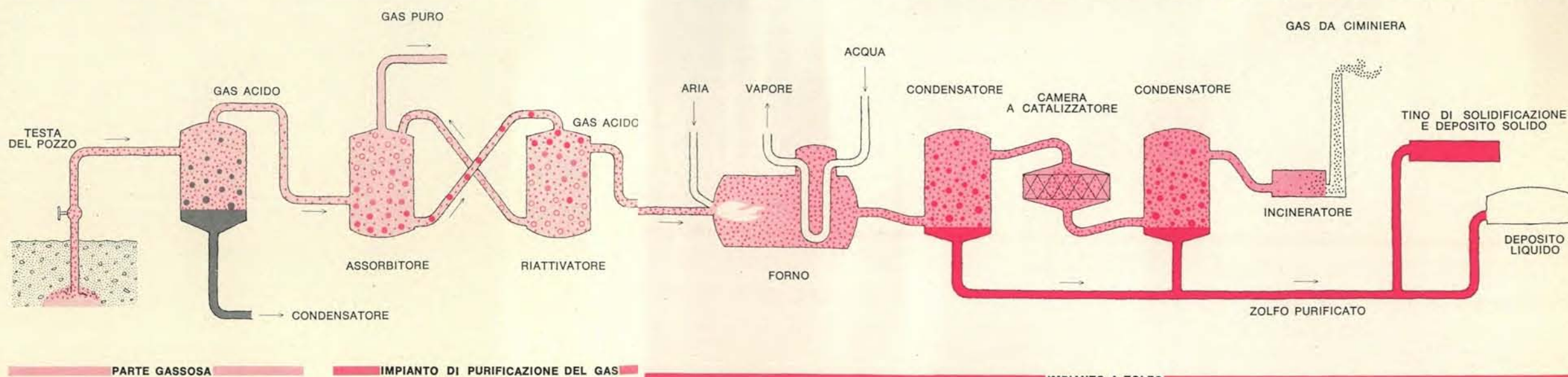
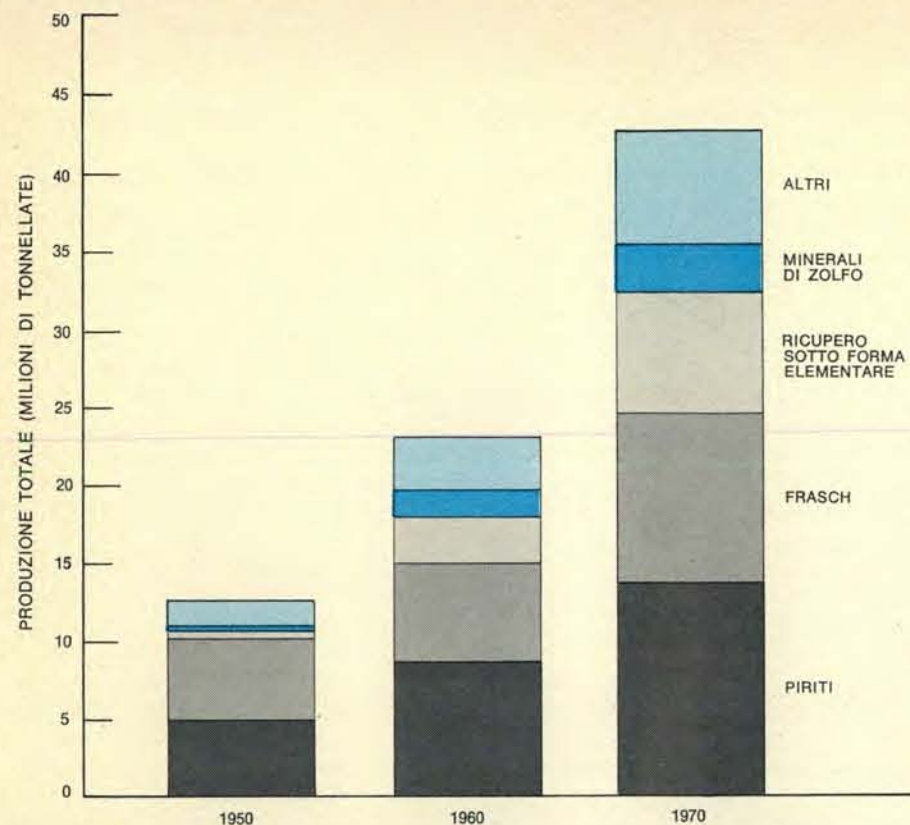
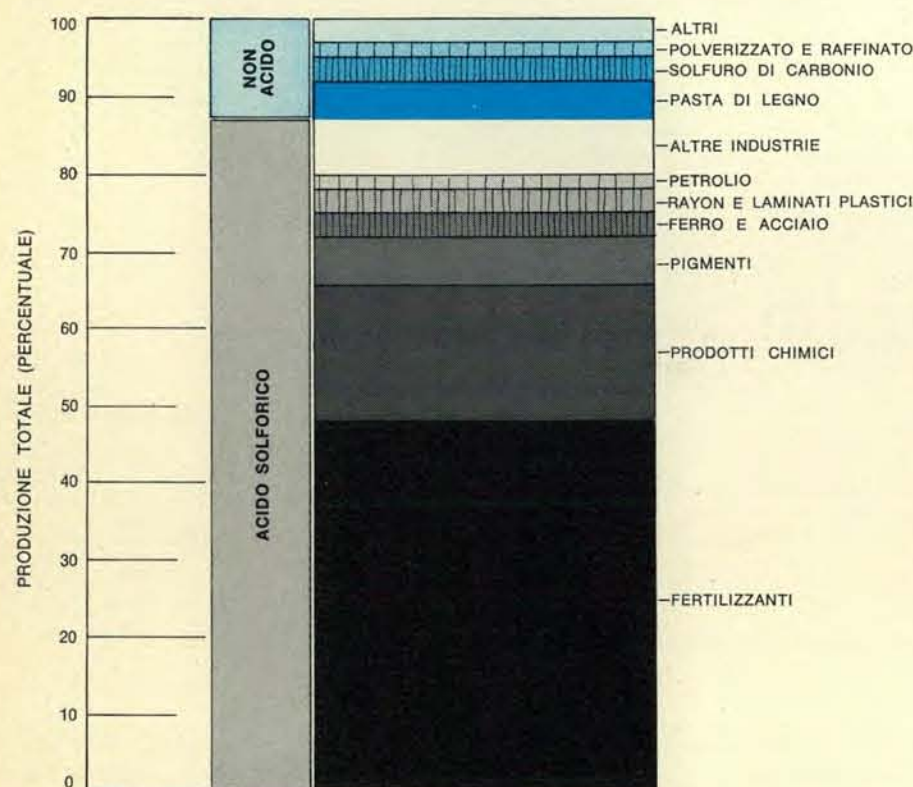
Acqua surriscaldata e ad alta pressione, con temperatura ben sopra il suo punto di ebollizione, è pompata nella intercapedine tra il tubo da 8 e quello da 4 pollici e anche, almeno all'inizio, in quello da 4 pollici. L'acqua surriscaldata esce dai fori della tubazione da 8 pollici nel deposito di zolfo. Lo zolfo fonde ed essendo più pesante dell'acqua si raccoglie in polla al fondo del pozzo. Si smette allora di pompare acqua nella tubazione da 4 pollici: la pressione statica dell'acqua, compressa nella formazione rocciosa, spinge lo zolfo liquido su per la tubazione da 4 pollici per qualche centinaio di metri. Dell'aria compressa, immessa nella tubazione da 1 pollice, forma una schiuma leggera di zolfo e permette di portarlo in superficie.

In superficie lo zolfo è raccolto in enormi tini, ove solidifica, oppure in recipienti riscaldati a vapore per la successiva spedizione su battelli riscaldati appositamente, o su carri merci o camion appositi che lo trasportano, per la vendita, allo stato fuso.

Un tino a zolfo, come quello riportato nella figura a pagina 39, è spettacolare. È costituito praticamente da un gran blocco giallo di zolfo solidificato, che può esser alto anche 15 metri, lungo 300 e largo 50 e contenere qualcosa come 400 000 tonnellate di zolfo. Un tino si forma spruzzando dello zolfo fuso sulla sua superficie. Delle forme di alluminio, incastrate una nell'altra e contornanti la parte superiore del tino, contengono lo zolfo fino a che si solidifica: quindi si smontano e rimontano più in alto per lo strato successivo. Lo zolfo si stacca dal tino per quel tanto che occorre per una certa spedizione per ferrovia, con chiatte o battelli.

Altri procedimenti

Grandi quantità di zolfo non elementare si trovano anche in altre forme. Queste riguardano i solfuri di ferro (piriti), i solfuri dei metalli pesanti non ferrosi e, in minor quantità, i solfati di calcio. Questi minerali sono particolarmente importanti nei paesi che non hanno giacimenti di zolfo elementare, come gran parte dell'Europa, del Giappone e di altre regioni. Solo in pochi casi si tenta di recuperare lo zolfo da questi minerali, in quanto che il costo dell'operazione è proibitivo. Essi vengono, invece, arrostiti in appositi forni, venendo così trasformati in anidride solforosa che a sua volta è convertita in acido solforico. Nel caso delle piriti l'ossido di ferro che rimane è spesso eliminato come scoria, a meno che non contenga altri elementi importanti co-

GAS PURO
GAS ACIDO
TESTA DEL POZZO
GAS ACIDO
ASSORBITORE
RIATTIVATORE
CONDENSATORE
ARIA
VAPORE
ACQUA
FORNO
CONDENSATORE
CAMERA A CATALIZZATORE
CONDENSATORE
GAS DA CIMINIERA
INCINERATORE
TINO DI SOLIDIFICAZIONE E DEPOSITO SOLIDO
DEPOSITO LIQUIDO
ZOLFO PURIFICATO
PARTE GASSOSA
IMPIANTO DI PURIFICAZIONE DEL GAS
IMPIANTO A ZOLFO

Il ricupero dello zolfo dai gas naturali acidi, chiamati così per il loro contenuto in idrogeno solforato, è diventata una delle più importanti fonti di zolfo. Nella prima fase del processo si separano gli idrocarburi liquidi dal gas in un opportuno condensatore. Questi vanno avviati in raffineria, mentre il gas, ancora acido, va a un impianto di purificazione in cui l'idrogeno solforato è assorbito da una soluzione amminica. Il gas purificato viene raccolto per la vendita. La soluzione in cui è stato assorbito l'idrogeno solforato, si porta invece in un riattivatore dove si libera l'idrogeno solforato, cioè il gas acido, che viene quindi portato a un impianto a zolfo. In questo impianto il gas si mescola ad aria e viene bruciato allo scopo di formare miscela di vapore e gas solforosi. La maggior parte dello zolfo viene condensata a questo punto: il gas residuo viene quindi condotto su un catalizzatore dove dell'altro idrogeno solforato si trasforma in vapore di zolfo. Nelle camere a condensazione lo zolfo diventa liquido. Il gas residuo viene poi alla fine bruciato.

La produzione dello zolfo è circa triplicata dal 1950 mentre è variata la proporzione delle varie fonti da cui lo si estrae. Le piriti sono minerali che contengono ferro e zolfo combinati. La maggior parte dell'aumento indicato con «altri» è dovuto al ricupero dello zolfo dai gas naturali acidi, dovuto alla richiesta di gas naturali puri.

Lo zolfo è usato soprattutto sotto forma di acido solforico. La sua importanza, nella produzione di fertilizzanti, consiste nella solubilizzazione delle rocce fosfatiche per ottenere fertilizzanti fosfatici. La maggior parte dello zolfo non acido è adoperata nell'industria della carta per sciogliere le impurezze della pasta di legno e sbiancarla. Zolfo in polvere è adoperato nella vulcanizzazione della gomma. Zolfo puro in medicina.

me il rame, in quantità sufficiente a giustificarne il recupero. I minerali non ferrosi sono adoperati soprattutto per il loro contenuto metallico. In questi casi l'anidride solforosa è un sottoprodotto, spesso non adoperato soprattutto se nelle vicinanze dell'impianto non esiste la possibilità di trasformarla in acido solforico: cosí viene dispersa nell'aria e costituisce uno dei componenti dell'inquinamento atmosferico.

Ci si può attendere che le richieste di una atmosfera piú pulita riducano lo scarico di fumi solforosi non solo da parte delle fonderie, ma anche di fabbriche e centrali elettriche che bruciano carbone o nafta ad alto tenore di zolfo. Altre fonti di zolfo, di secondaria importanza, riguardano i numerosi impianti che si stanno costruendo per la eliminazione e ricupero dello zolfo, in forma elementare, dalle frazioni pesanti della distillazione del petrolio.

Processi di eliminazione dello zolfo dai gas naturali acidi, chiamati acidi per la presenza di idrogeno solforato in alto contenuto, stanno diventando la fonte principale di zolfo per il Canada, l'Iran, la Francia, la Germania e altri paesi, man mano che cresce, corrispondentemente, il bisogno di gas naturali puri e di loro prodotti. Il primo passo in questi processi è di solito la rimozione di idrocarburi liquidi, che sono inviati in raffineria. Il gas è quindi seccato e avviato in impianti di purificazione, dove l'idrogeno solforato si elimina per assorbimento su soluzioni opportune. Questo passo permette di ottenere gas naturali puri, inviati in opportune condutture per la distribuzione. La soluzione che ha assorbito l'idrogeno solforato è trasportata a un riattivatore che permette la separazione dell'idrogeno solforato dalla soluzione, tramite vapore, e la sua successiva combustione con aria a dare zolfo liquido e prodotti solforosi gassosi. Una successiva riciclazione permette di estrarre altro zolfo liquido dai gas.

La produzione canadese di zolfo dai gas naturali acidi, si sta avvicinando a 4 milioni di tonnellate all'anno. Potrà raggiungere gli 8 milioni di tonnellate nel 1980, data la crescente richiesta di gas naturali puri nel Canada e nel nord degli USA.

Numerose industrie europee producono acido solforico per arrostimento di solfato di calcio con carbone e sabbia. Il procedimento permette di ottenere anche una uguale quantità di cemento clinker, usato per produrre il cemento portland. Un impianto nel Texas ha cominciato a produrre zolfo elementare per riduzione del gesso con l'idrogeno contenuto nei gas naturali.

Si prevede anche lo sviluppo di processi di estrazione dello zolfo da minerali sulfurei e dall'acqua di mare (utilizzando le incrostazioni delle caldaie dei grandi impianti di dissalazione). Un'altra possibilità riguarda nuove tecnologie che permettano di accelerare l'azione di certi microorganismi su minerali solfati e solfuri, azione che in condizioni normali richiede qualche milione di anni per trasformare questi minerali in giacimenti di qualche importanza.

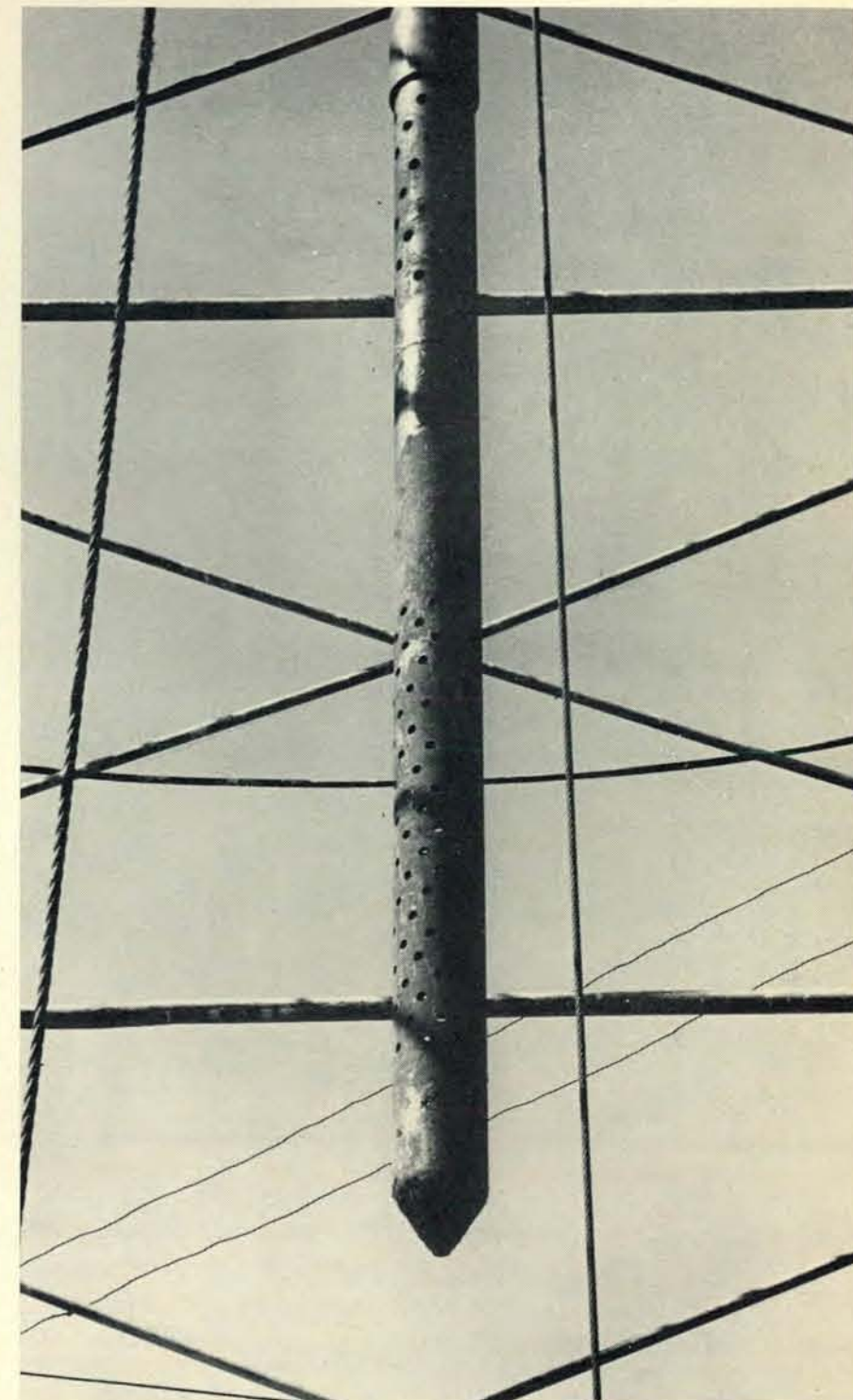
Usi dello zolfo

Poiché la maggior parte dello zolfo è trasformata in acido solforico e il 60 % di quest'ultimo è usato per la produzione di fertilizzanti, si può ben dire che circa la metà dello zolfo prodotto è usato in agricoltura. La maggior parte dell'acido solforico è usata per attaccare le rocce fosfatiche per produrre fertilizzanti come i perfosfati e il fosfato ammonico. In agricoltura si adoperano anche discrete quantità di solfato d'ammonio, che contiene sia lo zolfo che l'azoto.

Gli altri usi dell'acido solforico sono cosí numerosi che si può accennare solo ai principali. È usato nelle raffinerie per eliminare prodotti indesiderabili: nell'industria metallurgica per attaccare minerali come la bauxite, per produrre allumina e alluminio; serve ad attaccare la fluorite per produrre acido fluoridrico e la ilmenite per ottenere il biossido di titanio da usare come pigmento.

È meglio dividere i composti dello zolfo, oltre l'acido solforico, in derivati organici e inorganici. Strettamente connessi con l'acido solforico, nella categoria inorganica, sono l'acido solforoso e l'oleum. L'acido solforoso si prepara per assorbimento dell'anidride solforosa nell'acqua e l'oleum per assorbimento della anidride solforica nell'acido solforico. Entrambi i prodotti sono usati nell'industria chimica, che si occupa anche di altri composti dello zolfo con il cloro, il carbonio, il fosforo e l'idrogeno. Questi composti possono esser preparati riscaldando direttamente lo zolfo con l'altro elemento, spesso in presenza di un catalizzatore. Il monocloruro serve per vulcanizzare la gomma: il solfuro di carbonio è un solvente di odore sgradevole, il pentasolfuro di fosforo è un importante intermedio per la preparazione di certi insetticidi. Fra i composti organici dello zolfo ci sono i sulfamidici, i coloranti allo zolfo e i composti organici solfonati che servono da detergenti.

Prendendo in esame la sola economia degli USA, circa 85 % dello zolfo prodotto è usato come acido solforico. Esso si distribuisce per il 50% per i fertilizzanti, 20 % per i prodotti chimici, 4 % pigmenti e vernici, 4 % per le raffinerie di petrolio, 3 % per il rayon e fogli di plastica, 1 % ferro e acciaio, 3 % per le altre industrie. Il 15 % dello zolfo usato allo stato elementare è distribuito come segue: 6 % per l'industria della carta, 3 % per il solfuro di carbonio (adoperato per fabbricare il rayon, il cellofan e, tra l'altro, il tetracloruro di carbonio e come solvente della gomma); il 3% allo stato polverulento serve nella fabbricazione della gomma e puro a scopo medico: finalmente il 3 % è destinato ad altri scopi.

A parte la sua importanza industriale, lo zolfo svolge un ruolo fondamentale nei processi vitali. Se un terreno è

Tubazione d'acciaio, impiegata nel processo Frasch, pronta per esser adoperata in un pozzo. Quando si trova nel suo alloggiamento l'acqua calda esce dai fori superiori, nelle rocce. Lo zolfo fuso scorre dalle rocce nella tubazione attraverso i fori inferiori.

scarso in ioni solfato, lo sviluppo delle piante si arresta anche in abbondante presenza di altre sostanze nutritive. Questa condizione di deficienza la si ritrova come conseguenza di raccolti continuati o di scarso uso di fertilizzanti contenenti zolfo. Viene allora disturbata la formazione di certi composti organici contenenti zolfo, come gli amminoacidi metionina e cisteina, costituenti delle proteine vegetali. La deficienza di questi composti nella nutrizione delle piante, causa seri problemi perché si arrestano dei processi metabolici che portano alla formazione di enzimi, ormoni, pelle e peli.

Scorte e richieste di zolfo

L'accrescimento della popolazione e l'innalzamento del tenore di vita, hanno assieme prodotto un aumento nella domanda di zolfo che risulta, in media, del 4-5 % all'anno da parecchi anni. In altri termini, il consumo di zolfo è aumentato da circa un milione di tonnellate nel 1900 a circa 40 milioni di tonnellate nel 1969. Di conseguenza le scorte principali di zolfo si sono estese dai primi giacimenti in Sicilia, alle piriti, ai depositi Frasch, ai gas naturali e ad altre fonti ancora. Anche se la produzione è continuamente aumentata, la domanda ha superato l'offerta all'inizio degli anni 50 e verso la metà degli anni 60. In entrambe queste occasioni la scarsità di materia prima e gli alti prezzi hanno stimolato la ricerca mondiale di nuove fonti e portato a nuove tecnologie non solo da parte dei produttori di zolfo ma anche da parte dei consumatori che hanno cercato di ridurre la loro dipendenza dallo zolfo come materia prima. Per esempio alcune industrie europee di fertilizzanti fosfatici hanno cominciato a impiegare acido nitrico al posto dell'acido solforico e i produttori di acciaio sono passati all'acido cloridrico per la finitura.

Impianto a gas acido nella Germania occidentale: tratta l'idrogeno solforato che accompagna i vari gas naturali. L'idrogeno solforato è parzialmente bruciato ad anidride solforosa nella camera orizzontale che si vede a destra. I due gas reagiscono e sono quindi raffreddati a formare zolfo fuso che si raccoglie nella cavità a sinistra in primo piano. Questo procedimento permette di ricavare circa il 70 % dello zolfo presente nel gas. Il rimanente si ottiene per riscaldamento del gas residuo, che passa poi attraverso un catalizzatore a ossido di alluminio e lo zolfo viene ricuperato nel condensatore verticale a sinistra.

La combinazione di nuove fonti e di nuove tecnologie e il passaggio ad altre sostanze, ha portato alla fine della scarsità di zolfo verso la metà del 1968. Le nuove scorte dovrebbero bastare per una decina di anni. In piú si prospettano nuove fonti di zolfo, basate soprattutto sullo sforzo di ridurre l'inquinamento atmosferico. L'industria petrolifera sta affrontando la richiesta di oli combustibili a tenore di zolfo dell'1% e meno, e questo in svariati paesi. La rimozione dello zolfo da questi combustibili può costituire una delle piú importanti fonti di zolfo elementare, soprattutto se diventerà sempre piú in uso lo sfruttamento del petrolio proveniente da scisti o sabbie bituminose.

Il carbone ad alto contenuto in zolfo non viene trattato, al presente, prima della combustione con la stessa efficienza con cui sono trattati il petrolio e i gas naturali acidi. Questo perché lo zolfo è presente nel carbone combinato o sotto forma organica o di pirite. Non sono state ancora messe a punto opportune tecniche di recupero dello zolfo da questi prodotti. Si sta cercando di rimuovere lo zolfo dai gas di combustione di questi carboni, trasformandolo in acido solforico, solfato d'ammonio o zolfo elementare. Si reputa che solo negli USA lo zolfo disperso nell'atmosfera equivalga a qualcosa come 6 milioni di tonnellate all'anno. Anche se è probabile che il trattamento dei gas di ciminiera verrà presto fatto su grande scala, solo una parte dello zolfo potrà essere ricuperata, dati i costi del processo. Un'alternativa piú economica consiste nel trattare questi gas con calce ed eliminare i prodotti ottenuti. In ogni modo le scorte di zolfo sembrano notevoli. Si sta già presentando l'eventualità di una produzione in eccesso in Canada dovuta alla crescente domanda di gas naturali puri.

Sembra non potrà esserci scarsità di zolfo in un futuro anche remoto, soprattutto se la tecnologia permetterà di estrarlo economicamente dai miliardi di tonnellate esistenti di solfato di calcio, di carbone e di acqua di mare.

Comunicazione intercellulare

Le cellule dei tessuti viventi agiscono coordinatamente in vari modi. Un meccanismo osservato di recente permette il passaggio di molecole, che funzionano da segnali, da una cellula all'altra

di Werner R. Loewenstein

Un organismo multicellulare è costituito da una comunità di cellule che agiscono coordinatamente e sono disposte secondo uno schema ben preciso. Questa osservazione sembra ovvia, ma quando ci domandiamo in che modo questo ordine viene raggiunto ci troviamo immediatamente di fronte a uno dei misteri della vita. L'ordine comporta un qualche tipo di comunicazione, alcune delle quali sono conosciute piuttosto bene; per esempio, un ormone porta un messaggio a una cellula che è specificamente sensibile ad esso e un anticorpo riconosce un antigene della superficie cellulare dalla forma della sua molecola. Recentemente è stata trovata una forma di comunicazione in cui le parti interne delle cellule sono direttamente connesse mediante una particolare disposizione molecolare delle membrane della superficie cellulare. Questo tipo di comunicazione sembra particolarmente adatto per controllare il comportamento delle cellule in un organismo che si sviluppa.

a

INTERNO DELLA CELLULA

MEMBRANA DELLA CELLULA I

MEMBRANA DELLA CELLULA II

INTERNO DELLA CELLULA

b

INTERNO DELLA CELLULA

MEMBRANA DELLA CELLULA I

MEMBRANA DELLA CELLULA II

INTERNO DELLA CELLULA

La fotografia al microscopio elettronico della pagina a fronte, mostra le zone di contatto tra le membrane di due cellule epatiche: ingrandimento di 140 000 volte. S. Bullivant della University of Auckland ha preparato il campione per « rottura per congelamento » e l'ha ombreggiato con platino per mettere in evidenza i dettagli tridimensionali. I disegni (*sopra*) mostrano una sezione schematica trasversale del campione prima (*a*) e dopo (*b*) la rottura lungo il piano di contatto. La zona colorata rappresenta lo spazio intercellulare. Nella microfotografia si può vedere solo il frammento inferiore: la membrana superiore è stata strappata via tranne una parte sul piano di frattura (*a sinistra*). Nella foto la frattura si vede come una luminosa « scarpata » verticale. A sinistra di questa, dall'interno della cellula, si vede la parte interna della membrana della cellula I; i « rilievi » ravvicinati indicano una zona di intimo contatto con la membrana della cellula II. A destra, dallo spazio intercellulare, si vede la parte interna della membrana della cellula II. Il gruppo ovale (*estrema destra*) è una zona prima in intimo contatto con la membrana della cellula I che sta sopra. Nelle zone di contatto si possono formare giunzioni per le comunicazioni intercellulari.

La mia storia comincia con un'osservazione fatta per caso. Nel 1962 Yoshinobu Kanno e io stavamo studiando la permeabilità della membrana che avvolge il nucleo cellulare. Per questi esperimenti usavamo le grandi cellule delle ghiandole salivari della larva del moscerino della frutta. Iniettammo ioni nel nucleo di una cellula mediante una micropipetta producendo una corrente elettrica e scoprimmo con nostra grande sorpresa che il potenziale elettrico nella cellula adiacente era elevato quasi come quello nella cellula iniettata. Evidentemente una grande parte della corrente era passata da una cellula all'altra attraverso le regioni della membrana cellulare in cui le cellule erano a contatto. Piú o meno contemporaneamente Stephen W. Kuffler e David D. Potter alla Harvard Medical School fecero un'osservazione simile su cellule nervose satelliti della sanguisuga. Mezzo anno piú tardi Kanno e io iniettammo molecole relativamente grandi dentro una cellula delle ghiandole salivari e trovammo che anche esse passavano attraverso le regioni in cui le cellule erano a contatto.

Erano questi risultati sorprendenti. Era opinione comune che le cellule fossero unità separate e questo principio costituiva il nucleo della teoria cellulare, che era stata per buoni motivi fortemente sostenuta dai biologi. La teoria cellulare ebbe origine nella prima metà del XIX secolo, quando era stato osservato che le cellule erano unità anatomiche autoreplicantesi. Se qualche dubbio era ancora rimasto circa l'indipendenza anatomica delle cellule, questo fu dissipato in questi ultimi due

Questa foto coglie il momento della giunzione cellulare in cui due cellule isolate da un embrione di tritone sono messe a contatto per manipolazione. Il diametro delle cellule è di 0,3 millimetri; le misure dimostrano che esse comunicano in pochi secondi.

Si prepara la ghiandola salivare di un moscerino vivo per studi di comunicazione intercellulare inserendo micropipette in singole cellule, come si vede nella seconda cellula da sinistra. La tecnica usata per l'esperimento è illustrata nella pagina accanto.

decenni, quando il microscopio elettronico aveva rivelato una membrana continua che rivestiva tutte le cellule, anche là dove al microscopio ottico non si erano viste membrane. Si supponeva che questa membrana fosse una barriera continua; al microscopio elettronico furono osservate immagini analoghe in cellule di muscoli scheletrici, cellule nervose e globuli rossi e si era definitivamente dimostrato mediante tecniche osmotiche ed elettriche che i contorni di queste cellule costituivano una barriera continua. La membrana di queste cellule divenne cosí il prototipo della membrana superficiale della cellula. Le sole eccezioni erano le cellule cardiache e alcune fibre nervose che Silvio Weidman (che lavorava nel 1952 alla Università di Cambridge) ed Edwin J. Furshpan e Potter (che nel 1957 lavoravano all'University College di Londra) trovarono essere elettricamente connesse. Però, poiché queste cellule si specializzano a comunicare mediante impulsi elettrici, tali fenomeni erano propriamente considerati come specializzazioni adatte a questo tipo di comunicazione.

Quello che si riteneva fosse il prototipo della membrana cellulare si rivelò invece essere un'eccezione. Il lavoro svolto in questi ultimi anni nel nostro laboratorio al Columbia University College of Physicians and Surgeons ha dimostrato che le cellule di una grande varietà di tessuti animali che non sono implicate in segnalazioni elettriche comunicano attraverso giunzioni permeabili della membrana. Abbiamo trovato che questo è vero per tutte le cellule epiteliali esaminate: cellule delle ghiandole salivari, fegato, rene, tiroide, pelle, vescica urinaria, stomaco, intestino ed epiteli sensitivi.

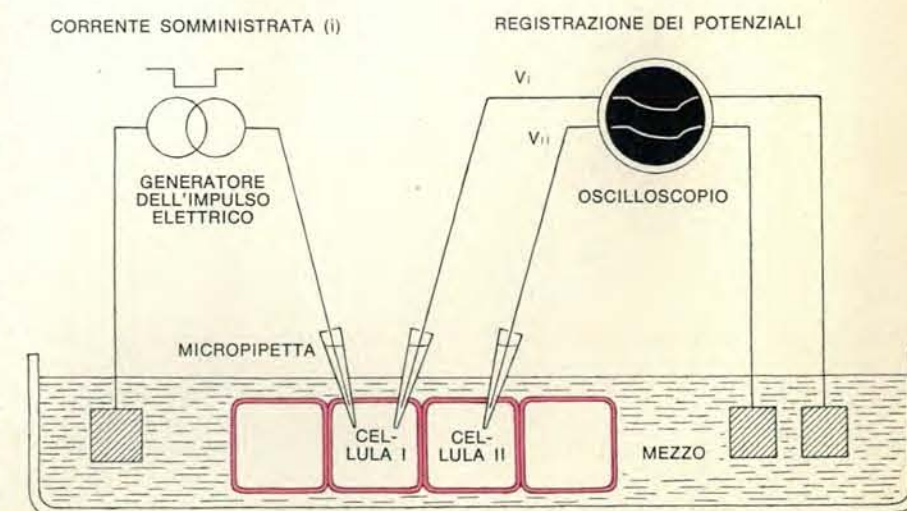
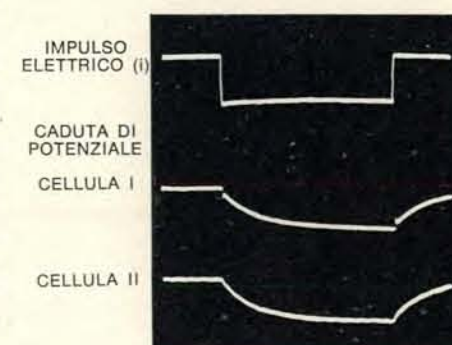
Misure elettriche eseguite su questi sistemi cellulari mostrarono che i piccoli ioni inorganici che portano le nostre correnti possono muoversi piuttosto liberamente dall'interno di una cellula all'interno di un'altra, ma non all'esterno. Laddove le cellule sono a contatto le regioni della membrana (membrane giunzionali) sono 1000 o 10 000 volte piú permeabili a questi ioni delle regioni della membrana affacciate all'esterno (membrane non giunzionali). Inoltre al livello della giunzione, l'interno del sistema è effettivamente isolato da quello esterno. Le misure elettriche cosí ci dicevano che queste giunzioni della membrana debbono essere organizzate in modo da lasciare dei passaggi tra le cellule. Ancora sappiamo molto poco sulla struttura di questi passaggi, ma possiamo avere un'idea generale sulla loro posizione nelle membrane cellulari. Sfruttando l'alta risoluzione del microscopio elettronico parecchi ricercatori hanno ottenuto fotografie che mostrano nelle zone di contatto tra membrane regioni di connessione organizzate. Sotto il nome di « giunzioni a contatto », « giunzioni distanziate » e « giunzioni settate » queste regioni (soprattutto le ultime due) sono probabilmente punti di comunicazione, perché esse mostrano una certa continuità strutturale tra le cellule e sono generalmente presenti in sistemi di cellule comunicanti. Però per risolvere il problema dell'organizzazione strutturale abbiamo bisogno di tecniche piú raffinate che non alterino la struttura della membrana come in genere capita con le tecniche di microscopia elettronica.

Poiché le cellule epiteliali normalmente non sono implicate in comunicazioni mediante segnali elettrici, sorse naturale la domanda se anche le particelle degli ioni potassio o cloro che hanno dimensioni maggiori e che portano le correnti elettriche dei nostri elettrodi, passino attraverso le giunzioni. Nel tentativo di rispondere a questa domanda usammo molecole fluorescenti o colorate in modo da poterle identificare dentro le cellule. Abbiamo iniettato queste molecole con micropipette sotto pressione idraulica o elettrica. La nostra prima molecola marcata è stata la fluoresceina che dà intensa fluorescenza quando è eccitata da radiazioni ultraviolette o dalla luce blu. Quando la fluoresceina veniva iniettata in una cellula di una ghiandola salivare e il sistema cellulare veniva osservato al microscopio sotto l'azione dell'illuminazione eccitante, osservavamo che le molecole passavano dall'interno di una cellula all'interno di un'altra senza perdite significative al di fuori delle cellule o delle loro giunzioni. Inoltre, il passaggio di fluoresceina veniva bloccato da trattamenti che impediscono il passaggio dei piccoli ioni; questo ci porta a credere che le particelle relativamente grandi di fluoresceina seguano la stessa via degli ioni piú piccoli attraverso le giunzioni cellulari.

Piú recentemente il passaggio di fluoresceina è stato anche dimostrato con esperimenti su cellule fibroblastiche eseguiti da Furshpan, Potter ed Edwin S. Lennox alla Harvard University e al Salk Institute for Biological Studies, e con esperimenti su una giunzione nervosa a trasmissione elettrica fatti da Michael van der Laan Bennet al College of Physicians and Surgeons. In ulteriori studi fatti nel nostro laboratorio abbiamo trovato che molecole con peso molecolare dell'ordine di 1000, e in alcuni casi dell'ordine di 10 000 (ma non di piú), passano attraverso le giunzioni cellulari nelle cellule delle ghiandole salivari. C'è spazio sufficiente per passaggi intercellulari di una grande varietà di sostanze.

Come si formano queste giunzioni? Il primo indizio ci viene dagli esperimenti che Shizuo Ito e io abbiamo fatto su cellule singole che mettevamo a contatto mentre registravamo la resistenza elettrica delle loro membrane superficiali. Cellule cosí accoppiate producevano una giunzione di comunicazione in un tempo sorprendentemente breve: in pochi minuti nel caso di cellule di spugne e in pochi secondi nel caso di cellule di giovani embrioni di tritone. Quando le cellule venivano di nuovo separate mediante manipolazione o trattamento chimico le membrane giunzionali, prima permeabili, si saldavano rapidamente e diventavano impermeabili come la membrana di una normale cellula libera. Una nuova giunzione di comunicazione si poteva poi ottenere mettendo le cellule di nuovo in contatto. Poiché il contatto avveniva a caso in differenti punti delle membrane la gran parte della membrana superficiale della cellula deve avere la potenzialità di formare una giunzione di comunicazione.

Un secondo indizio ci venne dagli esperimenti fatti in collaborazione con Muhamed Nakas e Sidney J. Socolar: essi mostrarono che la permeabilità delle membrane giunzionali dipende in qualche modo dallo ione calcio. Normalmente la concentrazione di ioni calcio nel citoplasma della cellula è di parecchi ordini di grandezza piú bassa di quella al di fuori della cellula. Abbiamo iniettato ioni calcio nelle cellule delle ghiandole salivari dei moscerini e abbiamo trovato che la permeabilità delle membrane giunzionali diminuiva rapidamente. Numerosi altri agenti che elevavano il livello degli ioni calcio nel citoplasma portavano allo stesso risultato. In condizioni in cui il livello del calcio doveva essere simile sia al di dentro che al di fuori, le membrane giunzionali non si potevano piú distinguere, in quanto a permeabilità da quelle non giunzionali; la comunicazione veniva veramente interrotta.

Un altro indizio fu fornito da esperimenti che mostravano come la permeabilità della membrana giunzionale dipenda dall'isolamento della giunzione. In questi esperimenti ho formato coppie di cellule di spugne e ho variato le concentrazioni di tre fattori che promuovono i processi di adesione delle cellule: una sostanza che contiene glicoproteine che sta alla superficie delle cellule (scoperta di Tom D. Humphreys e Aron A. Moscona alla University of Chicago), calcio e magnesio nel mezzo. Quando le concentrazioni di questi fattori nel mezzo superavano certi valori si aveva un buon isolamento intorno alla giunzione e le membrane divenivano permeabili in quella regione. Quando il mezzo scarseggiava in questi fattori non c'era isolamento e le regioni giunzionali della membrana rimanevano impermeabili come quando le cellule erano completamente separate. Per di piú, quando riducevo l'isolamento di una data giunzione togliendo dal mezzo

Si dimostra la comunicazione intercellulare con il passaggio di piccoli ioni inorganici da una cellula alla sua vicina (*sopra*). Due micropipette contenenti una soluzione salina sono inserite nella cellula I (*a sinistra*) e una pipetta simile è introdotta nella II. Un impulso elettrico è applicato alla prima cellula e i cambiamenti di differenza di potenziale che ne risultano attraverso le membrane superficiali di ambedue le cellule sono registrati dalla deflessione di due raggi di un oscilloscopio (*in basso*). A comunicazione completa e con gli ioni che si muovono liberamente da una cellula all'altra, l'impulso che dà una caduta nella cellula I è appena riscontrabile nella II.

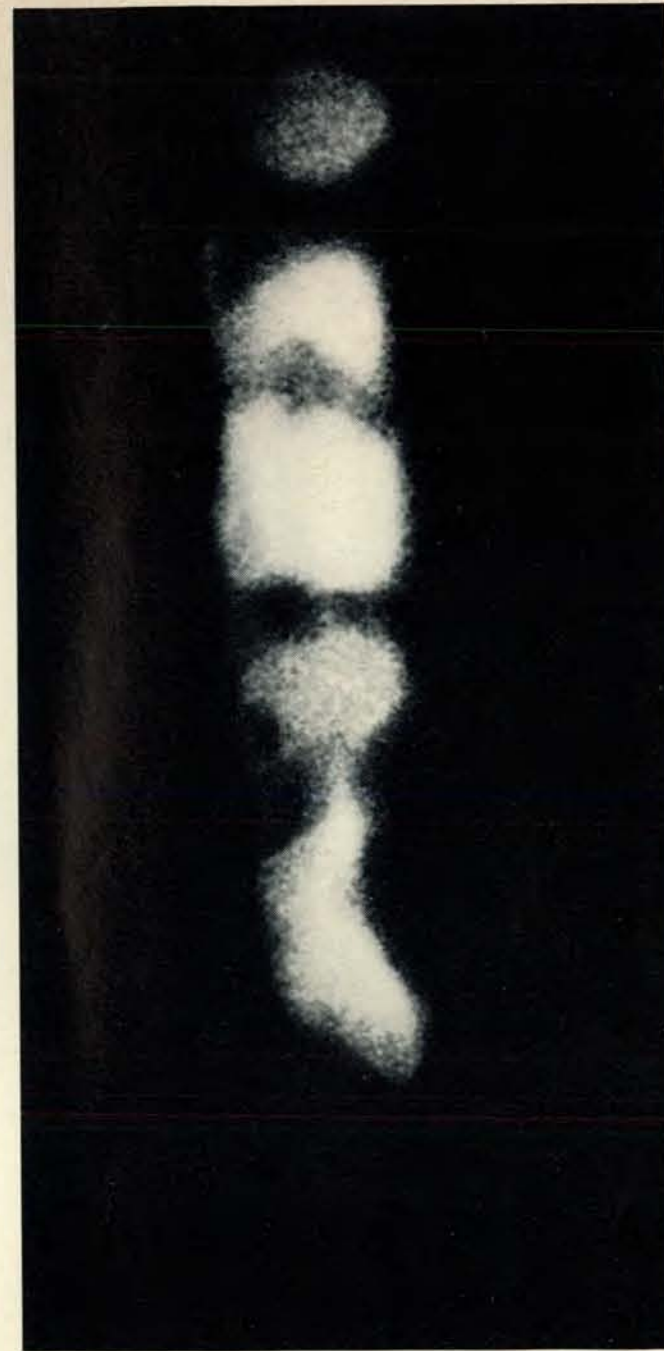
un po', ma non tutto, di calcio e di magnesio le membrane giunzionali, prima permeabili, si saldavano. Questo cambiamento da alta a bassa permeabilità fu dimostrato anche in parecchie cellule di animali superiori, comprese quelle dei mammiferi, piú organizzate e piú aderenti l'una all'altra; questo si ottenne trattando i tessuti con certi enzimi che abbassano l'isolamento.

Sulla base di questi risultati abbiamo formulato la semplice ipotesi che la comunicazione avvenga per un cambiamento di permeabilità delle membrane dovuto alla variazione della concentrazione di ioni calcio. Supponiamo prima di tutto che la diminuzione sperimentale della permeabilità di membrana con l'aumentare del calcio sia un aspetto di un processo a due direzioni; l'altro aspetto sarebbe l'aumento della permeabilità con il diminuire del calcio. Detta in altri termini, l'ipotesi è che certi siti leganti della membrana superficiale, occupati dal calcio allo stato impermeabile, siano liberi allo stato permeabile. Questa è una cosa ben nota per coloro che lavorano nel campo delle membrane biologiche; forse in questo caso l'unica caratteristica singolare è l'ampia variazione della permeabilità della membrana. Sulla base di questi nostri indizi sperimentali si può formulare la seguente ipotesi: nella formazione delle saldature isolanti intorno alle porzioni delle membrane cellulari a contatto, queste porzioni vengono incorporate nel compartimento intracellulare dove vengono a trovarsi in presenza di una concentrazione di ioni calcio di molti ordini di grandezza piú bassa rispetto a quando erano esposte al mezzo esterno. I gradienti chimici favoriscono ora il distacco del calcio da queste porzioni della membrana e ne risulta la loro conversione da bassa ad alta permeabilità (*si veda l'illustrazione in alto a pagina 54*).

Questa è stata la nostra ipotesi di lavoro per due anni. Tale ipotesi è attraente perché non richiede speciali proprietà delle regioni della membrana che stanno per congiungersi né meccanismi speciali per instaurare la comunicazione. L'elemento essenziale per iniziare la trasformazione di permeabilità e limitarla alle sole regioni giunzionali è la saldatura isolante, e le forze che guidano la trasformazione sono le stesse che mantengono basso il livello del calcio dentro le cellule. Il destino del calcio della membrana all'interno del sistema della cellula è lo stesso di quello di tutti gli ioni calcio del citoplasma. Il calcio va nei depositi cellulari, i mitocondri e altri organelli, oppure fuori attraverso le membrane non giunzionali. Quanto al destino dello ione calcio sul lato esterno delle membrane, esso non richiede una considerazione particolare. Se il microscopio elettronico dice il vero, non c'è posto per contenere una notevole quantità di ioni calcio tra le porzioni della membrana strettamente congiunte.

Molti e importanti aspetti dell'ipotesi formulata sulla funzione del calcio non sono ancora stati dimostrati. Non abbiamo nessuna prova diretta che la riduzione di permeabilità indotta dal calcio nelle membrane giunzionali sia reversibile o che lo ione calcio reagisca direttamente con la membrana. Comunque l'ipotesi ci ha portato a fare nuovi esperimenti che ci hanno detto qualcosa di piú sul meccanismo responsabile della comunicazione tra cellule.

L'idea centrale di questa ipotesi è che il basso livello degli ioni calcio all'interno della cellula sia responsabile dello stato di alta permeabilità nelle membrane giunzionali. È noto che il mantenimento del basso livello di calcio in parecchi tipi cellulari richiede un continuo trasporto di ioni verso il mezzo esterno ricco di calcio con spesa di energia fornita dal metabolismo cellulare. Ci si dovrebbe quindi aspettare che un'inibizione del metabolismo riduca la permeabilità delle membrane giunzionali. Questa considerazione suggerí ad Alberto Politoff, a Socolar e a me di studiare gli effetti di inibitori metabolici sulla comunicazione giunzionale in cellule di ghiandole salivari di moscerino. In un gruppo di esperimenti abbiamo inibito il metabolismo della cellula raffreddando il tessuto o aggiungendo al mezzo una piccola quantità di veleno (come cianuro, dinitrofenolo o oligomicina) che blocca la sintesi di adenosintrifosfato (ATP) e altre sorgenti di energia della cellula. Questo trattamento provoca una diminuzione della permeabilità delle membrane giunzionali interrompendo completamente la comunicazione tra cellule. Occorreva un po' di tempo, dai quindici minuti a due ore, perché questa interruzione avesse inizio. L'interruzione dovuta ad alcuni tipi di trattamento era reversibile; trattamenti ripetuti poi producevano interruzione senza praticamente alcun ritardo. Probabilmente con la ripetizione del trattamento, le cellule non avevano piú una riserva di ATP come invece al momento del primo trattamento. In un altro gruppo di esperimenti abbiamo combinato il trattamento di inibizione del metabolismo (dinitrofenolo) con la iniezione di ATP nelle cellule. Questo trattamento preveniva o invertiva l'interruzione in buona parte degli esperimenti (*si veda l'illustrazione in basso a pagina 54*).

Un'altra conseguenza dell'ipotesi del calcio è stata quella di avere trovato che la sostituzione del sodio con il litio nel mezzo extracellulare provoca interruzione. Si sapeva che questa sostituzione causava un aumento del calcio citoplasmatico in vari tessuti. Peter Baker, Mordecai Blaustein, Alan L. Hodgkin e Robert Steinhardt alla University of Cambridge hanno recentemente dimostrato che nelle fibre nervose della seppia un tale aumento era dovuto a un rallentamento del flusso di ioni calcio verso l'esterno e a un aumento di tale flusso verso l'interno. Nel nostro laboratorio Birgit Rose ha cominciato ora a studiare gli effetti della sostituzione del litio sulla comunicazione cellulare e ha trovato che in tre ore dalla sostituzione di tutto il sodio con il litio nel mezzo in cui sono immerse le ghiandole salivari di moscerino le cellule non comunicano piú. Che questo non fosse dovuto semplicemente alla mancanza di ioni sodio era dimostrato da esperimenti in cui il sodio veniva sostituito con la colina invece del litio; la comunicazione veniva mantenuta proprio come quando era presente il sodio.

A questo punto ci domandiamo che genere d'informazione possa passare attraverso le giunzioni cellulari. La cosa piú interessante a questo riguardo è che molecole piuttosto grandi possono passare per questa via da cellula a cellula. La grandezza di queste particelle che passano attraverso le giunzioni è abbastanza varia e comprende gran parte delle molecole implicate nel metabolismo (metaboliti) e molte altre molecole che regolano le attività cellulari. Cosí si presentò l'attraente possibilità che tra le molecole che potevano passare fossero incluse anche sostanze che regolano l'attività genica o, a rischio di osare troppo, che la giunzione fosse un mezzo per convogliare sostanze che controllano la crescita e il differenziamento delle cellule. Questa possibilità era attraente perché gli studi di genetica microbiologica dei passati dieci anni hanno dimostrato come i metaboliti e altre molecole di misura analoga possano regolare i geni nei batteri, e perché gli studi di embriologia sperimentale degli ultimi quarant'anni ci hanno

Rappresentazione schematica delle unità di passaggio in una porzione di giunzione cellulare: sono schematizzati in alto (*a*) gli elementi di ciascuna unità identificati per mezzo di misure elettriche intracellulari, e in basso (*b*) una serie di unità tra due membrane cellulari in stretto contatto. I tre elementi costitutivi di un'unità sono la membrana non giunzionale (relativamente impermeabile), la membrana giunzionale (permeabile) e le « saldature » dove si incontrano le membrane non giunzionali di due cellule. (Le saldature possono essere formate solamente dal contatto tra due membrane e potrebbero presentare qualche tipo particolare di struttura.)

Dimostrazione del flusso da cellula a cellula di molecole relativamente grandi nelle ghiandole salivari di un moscerino. In una cellula iniettata con marcatore fluorescente (*freccia*) eccitato con luce blu, questo si diffonde nelle cellule vicine in 10 minuti.

L'ipotesi della conversione suppone che la permeabilità della membrana cellulare sia determinata dalla concentrazione di ioni calcio («a», *punti colorati*) nel mezzo situato ai due lati della membrana. Su una superficie della membrana esposta a un mezzo extracellulare normale, con una concentrazione di calcio maggiore di un millesimo di mole per litro, i siti leganti del calcio sono per la maggior parte occupati. All'interno della cellula gli ioni calcio sono più rari e vengono costantemente mandati fuori (*frecce ondulate*) dall'energia prodotta dal metabolismo cellulare e i siti leganti del calcio sulla membrana interna sono scarsamente occupati. Quando due cellule vengono a contatto (*b*), le saldature racchiudono regioni della superficie della membrana che prima erano esposte al mezzo cellulare. Queste regioni perdono i loro ioni (*frecce dritte*) e diventano permeabili (*area bianca*). Il processo inverso può essere dimostrato sperimentalmente (*c*). Le membrane giunzionali tra cellule comunicanti diventano impermeabili quando ioni calcio sono artificialmente iniettati all'interno di una cellula.

insegnato che il differenziamento embrionale comporta qualche forma di interazione a breve distanza tra cellule mediante molecole diffusibili. Le comunicazioni giunzionali rappresentavano un buon candidato per le interazioni a breve distanza: un sistema chiuso in cui le molecole possano diffondere da citoplasma a citoplasma con poca perdita. Naturalmente ci sono altre possibili forme di comunicazioni intercellulari a breve distanza; per esempio molecole possono passare da una cellula all'altra attraverso il liquido extracellulare, e questa forma di comunicazione è indubbiamente importante nello sviluppo embrionale. La comunicazione giunzionale sembra particolarmente adatta nei processi di sviluppo in cui deve essere trasmessa l'informazione circa il numero e la posizione delle cellule in una comunità cellulare. La comunità di cellule ha un volume finito e un confine periferico di diffusione ben determinato (le membrane non giunzionali e le saldature giunzionali), e cosí ha la possibilità di ottenere tale informazione sulla base di semplici cambiamenti della concentrazione di molecole diffusibili che funzionano da segnali.

Quando trovammo che una particella grande come la fluoresceina poteva passare attraverso una giunzione cellulare, ci mettemmo immediatamente a esaminare la possibilità che il sistema di comunicazione trasmetta molecole-segnali per la regolazione della crescita e del differenziamento cellulare. Ci sarebbe piaciuto avere a che fare con segnali di comunicazione, ma in nessun caso di sviluppo embrionale a breve distanza è stata identificata una molecola-segnale. Perciò abbiamo dovuto ricorrere alla tattica indiretta di studiare la comunicazione in cellule aberranti in cui il controllo della crescita è stato ovviamente alterato.

Abbiamo cominciato con esperimenti rivolti a determinare se le cellule in un organismo in crescita comunicano veramente per mezzo di giunzioni. I nostri primi tentativi in questa direzione furono scoraggianti. Robert Ashman, Kanno e io passammo l'estate del 1964 al Marine Biological Laboratory di Woods Hole studiando la comunicazione tra le cellule dell'embrione di stella di mare: le prime figlie della cellula uovo. In queste cellule grandi e trasparenti abbiamo potuto registrare elettricamente la comunicazione durante la divisione della cellula uovo, da quando la cellula comincia a segmentarsi fino alla formazione del primo paio di cellule dell'organismo. Abbiamo riscontrato una buona comunicazione all'inizio della segmentazione, quando c'era ancora probabilmente continuità protoplasmatica tra le cellule. Però quando la segmentazione era completata la coppia di cellule cessava di comunicare. Questo era l'opposto di quello che ci saremmo aspettato. Un altro tentativo fatto l'estate successiva da Ito e da me sul primo paio di cellule di un altro organismo marino, il riccio di mare, dette lo stesso deludente risultato.

Un anno dopo il mio collega Ito e anche Furshpan, Potter e Lennox all'Harvard ottennero risultati piú soddisfacenti con embrioni piú avanzati. Ito ottenne nei tritoni la prova che molte, e forse tutte, le cellule dell'embrione sono in comunicazione quando questo ha raggiunto lo stadio di morula in cui le cellule formano una sfera. Il gruppo di Harvard, lavorando sull'embrione di seppia e piú tardi, quando a esso si aggiunse Judson Sheridan, con l'embrione di pollo, trovò numerose comunicazioni giunzionali anche a stadi piú avanzati, quando le cellule si erano visibilmente differenziate. I risultati del gruppo di Harvard erano particolarmente interessanti; essi mostrarono che c'era comunicazione tra vari tipi di tessuti embrionali i cui discendenti cellulari a differenziamento raggiunto sono chiaramente isolati nell'animale adulto e che certe cellule perdono le loro connessioni nel corso dello sviluppo. Recentemente Bennet e J. P. Trinkhaus riscontrarono anche comunicazione in stadi embrionali precoci del pesce *Fundulus* e Christina Slack e J. F. Palmer al Middlesex Hospital di Londra nel rospo *Xenopus*. Le osservazioni su questi cinque animali appartenenti a gruppi molto differenti stabilirono che la comunicazione giunzionale è presente durante il periodo di attiva crescita e differenziamento.

I nostri studi sulle cellule aberranti hanno avuto origine dalla considerazione che se la giunzione fosse veramente una via d'ingresso per le sostanze che regolano la crescita, ci si dovrebbe aspettare che un'anormale crescita di cellule, come il cancro, potrebbe in certi casi essere dovuta a scarsa comunicazione giunzionale. Speravamo che le nostre tecniche elettriche per registrare il passaggio di piccoli ioni inorganici al livello della giunzione fossero utili per individuare qualche alterazione di questo tipo di comunicazione.

Un materiale adatto per questo studio è il fegato dei mammiferi. Il mio collega R. D. Penn aveva già trovato che le cellule di fegato normale hanno una buona comunicazione giunzionale e vari tipi di tumori epatici dei mammiferi, prodotti sperimentalmente, si potevano facilmente avere da parecchi laboratori. Kanno e io abbiamo esaminato quattro tipi di tali tumori, che ricoprivano un vasto spettro di velocità di crescita e di differenziamento; in nessuno abbiamo potuto trovare comunicazione. Esperimenti successivi provarono che questo era anche vero per tumori della tiroide nel ratto e dell'epitelio gastrico dell'uomo. I miei colleghi Carmia Borek e Shoji Higashino eseguirono esperimenti analoghi con cellule di fegato in coltura. In questi esperimenti cellule di fegato normale e canceroso venivano coltivate in recipienti di coltura di vetro dove era possibile eseguire misure elettriche della comunicazione in condizioni molto simili per tutti i tipi cellulari. Anche qui soltanto le giunzioni tra cellule normali risultavano comunicanti; non si riscontrava comunicazione tra una cellula cancerosa e un'altra o tra una cellula cancerosa e una normale. La membrana di queste cellule cancerose, contrariamente a quelle normali, sono chiaramente impermeabili persino dove le cellule sono a contatto. Le regioni di contatto appaiono anche differenti al microscopio elettronico; numerosi ricercatori hanno trovato che il fegato canceroso e certi altri tessuti cancerosi mancano di alcune strutture che normalmente tengono unite le cellule.

La mancanza delle comunicazioni giunzionali sembra essere una manifestazione della crescita cancerosa, ma non della crescita in generale. Penn e io abbiamo dimostrato questo in uno studio sul fegato rigenerante. Il fegato del ratto adulto, come il fegato del leggendario Prometeo, è capace di rigenerarsi rapidamente. Se una parte del fegato, diciamo i due terzi, viene asportata, il moncone cresce durante i primi quattro giorni al ritmo di un miliardo di cellule al giorno (un ritmo molto piú veloce di quello del tumore piú rapido da noi esaminato). Successivamente la crescita rallenta. Entro il settimo giorno, quando il fegato ha raggiunto circa il suo volume, la crescita si riduce a meno dell'8 % del suo massimo valore. L'andamento di questa crescita rigenerativa rispecchia quello della crescita normale in un organo embrionale. Si tratta di una popolazione cellulare che sa quando e dove smettere di crescere. Abbiamo infatti trovato che tale popolazione di cellule in rigenerazione possiede comunicazioni giunzionali.

Non tutti i tipi di cellule cancerose mancano di comunicazioni giunzionali. Durante gli ultimi quattro anni il gruppo di Harvard e il nostro hanno trovato un certo numero di cellule cancerose che non presentano difetti nel passaggio di piccoli ioni inorganici o (in un caso) nel passaggio di fluoresceina. Questo non era strano. Difetti nella comunicazione giunzionale non sono naturalmente la sola causa possibile della crescita non controllata. La crescita cancerosa può avere origine da difetti nella produzione o nella ricezione di sostanze che controllano la crescita, che non hanno a che vedere con i difetti del loro passaggio intercellulare. Però la scoperta che alcune forme di cancro sono associate con difetti nella comunicazione intercellulare è un passo incoraggiante. Essa apre un nuovo capitolo nella storia della crescita cellulare e la ricerca in questa direzione sta attualmente progredendo.

a

La funzione dell'energia metabolica per mantenere la permeabilità della membrana giunzionale è dimostrata nel caso di due cellule comunicanti (*a*). Al mezzo extracellulare viene aggiunto dinitrofenolo, che blocca la sintesi dell'adenosintrifosfato (ATP). Dopo quindici minuti la permeabilità della membrana giunzionale è diminuita (*b*); le cellule non comunicano piú. L'interruzione della comunicazione avviene probabilmente perché non c'è piú adenosintrifosfato disponibile per fornire l'energia necessaria a far uscire gli ioni calcio dall'interno della cellula. Quando questi ioni aumentano nel citoplasma la membrana diventa impermeabile (*si veda l'illustrazione precedente*). Poi (*c*) l'adenosintrifosfato viene iniettato in una cellula. L'eliminazione del calcio e la comunicazione riprendono nonostante la presenza del dinitrofenolo nel mezzo.

La comunicazione tra cellule embrionali è determinata elettricamente misurando la resistenza al passaggio di ioni attraverso il piano di segmentazione. Gli elettrodi sono inseriti (*in alto*) in un uovo fecondato di stella di mare. A divisione cellulare completa (*sotto*), la comunicazione tra le due cellule è quasi interrotta. Essa riprende in uno stadio piú avanzato.

Fuochi d'artificio nello spazio

Un nuovo metodo per la ricerca spaziale consiste nel generare nubi artificiali di plasma nello spazio allo scopo di indagare sui campi elettrici e magnetici ivi esistenti

di Cristiano Batalli Cosmovici

« D'un tratto si assisteva a una esplosione, come al nascere di una stella. Dopo qualche secondo la luminosità diminuiva e si delineava una forma circolare di color giallo e verde; il disco rimaneva immobile mutando il suo colore in azzurro scuro, mentre una parte si staccava assumendo una forma sigariforme e un colore purpureo. » Cosí, qualche anno fa, uno spettatore ignaro descriveva lo spettacolo strano e fantastico a cui aveva assistito osservando il cielo al tramonto.

Non si trattava di oggetti non identificati o dischi volanti, ma di un genere nuovo e spettacolare di esperimenti nello spazio, ideati e attuati da un gruppo di scienziati dell'Istituto Max-Planck per la fisica extraterrestre di Garching (Monaco di Baviera).

I metodi della ricerca spaziale hanno consentito negli ultimi anni di allargare in modo vertiginoso la conoscenza del cosmo e dei fenomeni fisici che in esso avvengono, consentendoci non solo di studiare lo stato attuale dell'universo, ma di risalire gradualmente, attraverso le osservazioni e i dati acquisiti, alla sua formazione e di « scriverne la storia ».

Le informazioni fornite, in quest'ultimo decennio, da palloni stratosferici, razzi, satelliti artificiali e sonde spaziali non potevano, però, aiutarci a risolvere tutti i problemi connessi con questa giovane scienza. Bisognava risolvere un problema molto importante: lo studio del comportamento di un plasma nello spazio, ove le condizioni ambientali sono difficilmente o affatto riproducibili in laboratorio.

Camera a vuoto di 50 m^3 (in cui si ottiene una pressione di 10^{-6} torr) per studiare le proprietà fisico-chimiche delle miscele esplosive. Per maneggiare il bario sono necessarie speciali misure di sicurezza: laboratori isolati dall'esterno e tute corazzate.

Il plasma nello spazio extraterrestre

Il plasma, o gas ionizzato, consiste di atomi o molecole carichi positivamente, detti ioni, e di elettroni liberi carichi negativamente. Sebbene costituisca circa il 90 % di tutta la materia cosmica, esso è molto rarefatto nello spazio interplanetario e interstellare ed è visibile solo nelle code delle comete e nella corona solare. Un plasma si comporta, a causa delle sue proprietà elettriche, in modo differente da un normale gas elettricamente neutro, perché le particelle cariche del plasma sono influenzate nel loro movimento da campi elettrici e magnetici. In un campo elettrico tali particelle si muovono lungo le linee di forza del campo, mentre in un campo magnetico descrivono un moto elicoidale intorno alle linee del campo (*si veda l'illustrazione a pagina 57 in alto*).

Nello spazio extraterrestre campi elettrici e magnetici agiscono contemporaneamente sul plasma e, sotto la risultante di questi campi, che per semplicità assumiamo perpendicolari fra di loro (*si veda l'illustrazione a pagina 57 in basso*), la particella carica, oltre che al movimento elicoidale intorno alle linee del campo magnetico, si muove in direzione perpendicolare sia al campo magnetico sia a quello elettrico; tale moto, detto di deriva (*drift*), è indipendente dal segno e dalla massa della particella, cioè gli ioni positivi e gli elettroni negativi percorrono questa traiettoria di deriva nella stessa direzione e con la stessa velocità.

In un plasma può naturalmente accadere che le particelle urtino fra di loro durante il loro moto intorno alle linee del campo e, di conseguenza, saltino lungo un'altra linea di forza dove continueranno a muoversi di moto elicoidale fino al successivo urto. Il plasma cosmico è però molto rarefatto e le collisioni fra le particelle sono molto rare. La densità del plasma nello spazio interplanetario è stata misurata da sonde spaziali e satelliti artificiali ed è dell'ordine di 1-10 particelle al centimetro cubo, mentre nello spazio interstellare è al massimo di una particella per centimetro cubo.

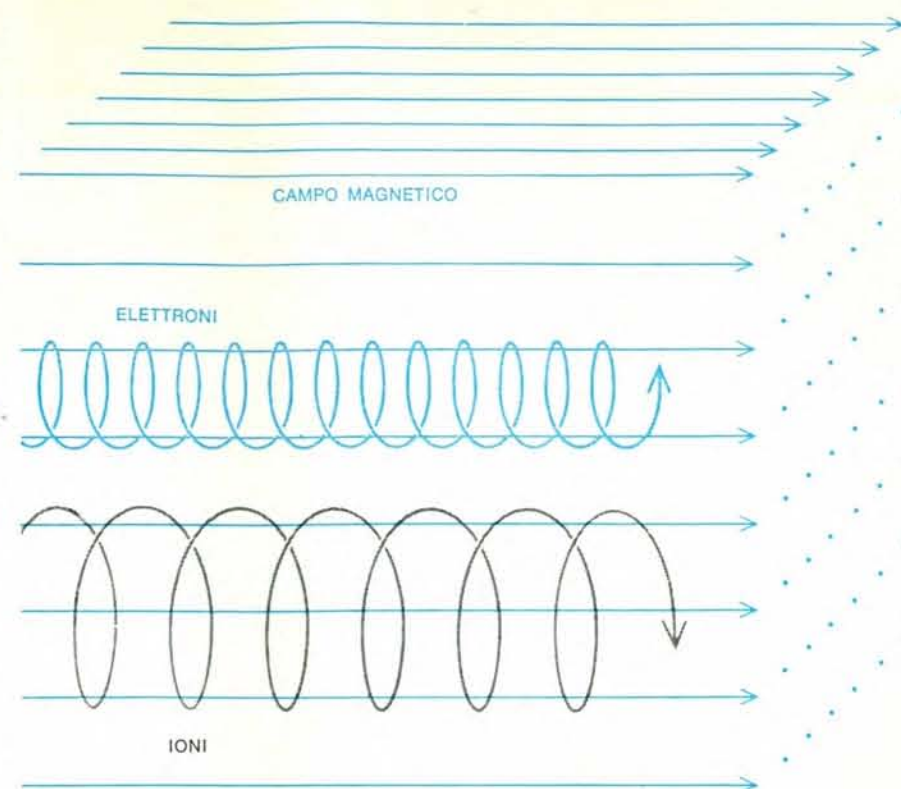
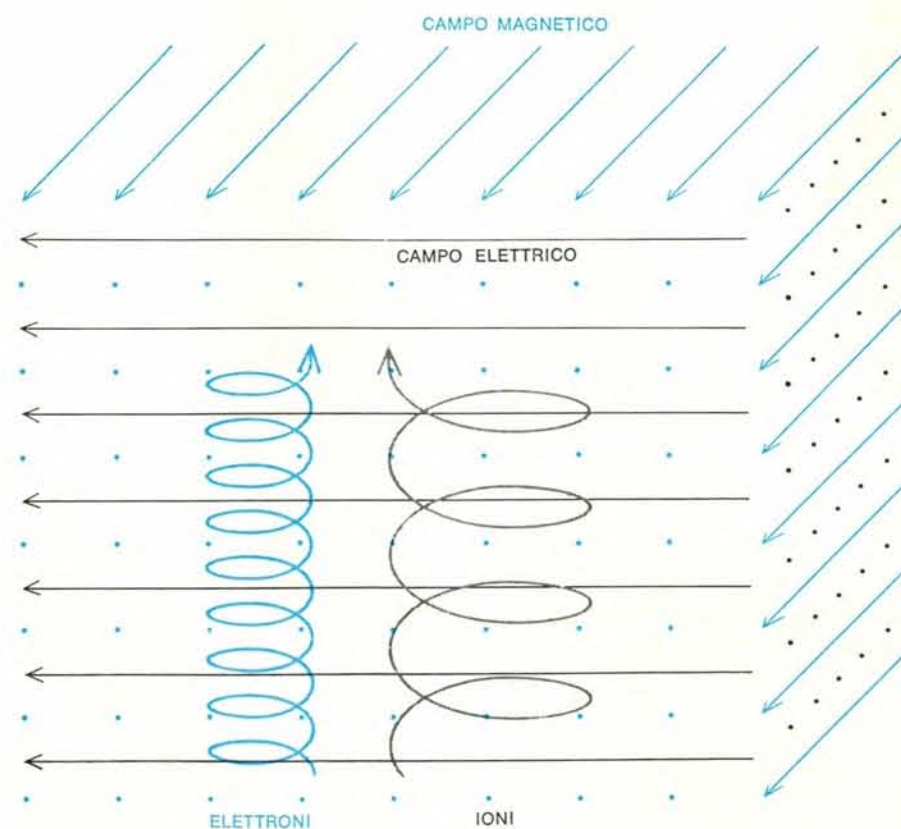
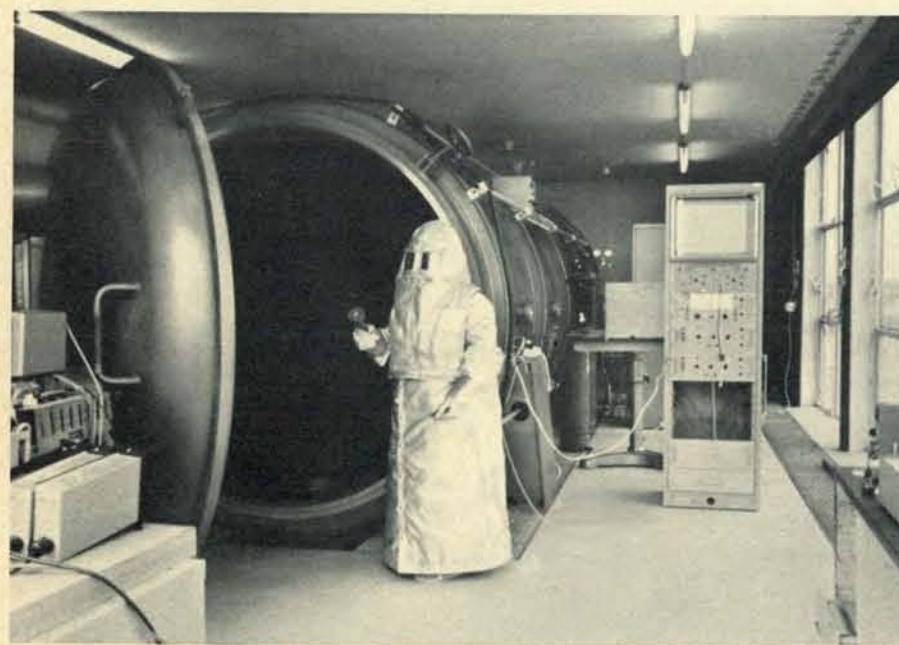
A una tale rarefazione, ioni ed elettroni urtano fra di loro ogni dieci ore circa. Il campo magnetico nello spazio interplanetario è quasi 100 000 volte piú debole che alla superficie terrestre (ove varia fra 0,3 e 0,6 gauss); in un tale campo i protoni impiegano circa due secondi per compiere un'orbita intorno alla linea del campo, mentre gli elettroni, data la loro minore massa, girano con una velocità duemila volte superiore. Comparando questi tempi, cioè queste frequenze di rotazione, con la probabilità d'urto, si riconosce facilmente che è lecito trascurare le collisioni nello spazio e supporre che le particelle eseguano indisturbate il loro moto di deriva.

A un osservatore si presenta allora il seguente quadro: le particelle rimangono « attaccate » alle linee del campo magnetico e il moto di deriva appare come un moto delle linee stesse del campo. Ne segue che noi potremmo vedere tale moto delle linee del campo magnetico nel momento in cui fosse visibile il plasma cosmico.

Il vento solare

Da uno studio approfondito sulle comete e dalla constatazione che le loro code si dispongono sempre in direzione opposta al Sole, si pensò all'esistenza di un flusso di particelle emesso in continuazione dal Sole: il vento solare. Ludwig Biermann postulò per primo nel 1951 tale esistenza e qualche anno dopo la sua teoria fu confermata sperimentalmente da misure eseguite per mezzo di sonde spaziali.

Per cercare di capire il complesso meccanismo di interazione fra questi plasmi (code di comete e vento solare) a noi non accessibili, Ludwig Biermann, Reimar Lüst e i loro collaboratori pensarono nel 1961 di generare nello spazio una nube di plasma artificiale, di cui fossero note la massa e la composizione chimica e che fosse osservabile da Terra. Inoltre tale nube poteva servire a studiare direttamente il vento solare e il campo magnetico nello spazio interplanetario, nonché la magnetosfera e la ionosfera terrestre. Mentre le particelle e i campi magnetici presenti intorno al nostro pianeta sono stati misurati e analizzati, non si conosceva nulla sui campi elettrici ivi esistenti.

Lo scopo degli esperimenti del grup-

Comportamento di un plasma in presenza di un campo magnetico. Il movimento elicoidale delle particelle intorno alle linee del campo magnetico è dovuto a due componenti: un movimento circolare intorno alle linee del campo e uno rettilineo nella direzione delle stesse: il primo è dovuto alla componente della velocità perpendicolare alle linee del campo, il secondo alla componente parallela di tale velocità.

Comportamento di un plasma in presenza di un campo magnetico e di un campo elettrico mutuamente perpendicolari. Le particelle eseguono in questo caso un movimento di deriva (*drift*) a sua volta perpendicolare a entrambi i campi. Osservando questo movimento e conoscendo il campo magnetico, si può risalire al campo elettrico.

po di Garching si può quindi considerare triplice: misurare i campi elettrici nello spazio extraterrestre; studiare il plasma rarefatto a una distanza di parecchi raggi terrestri (un raggio è di 6378 chilometri); preparare esperimenti con nubi di plasma nello spazio interplanetario.

Vediamo ora in che modo e con quali criteri bisogna procedere per generare nubi di plasma artificiale nello spazio.

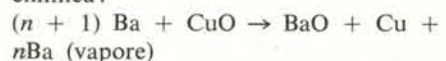
Nubi di plasma artificiale nello spazio

Sarebbe logico pensare di usare, per tali esperimenti, gli stessi atomi o molecole di cui è formato il plasma cosmico. Esso consiste per lo piú di idrogeno, ma questo elemento non è osservabile da Terra per il fatto che non diffonde sufficientemente la luce visibile. Nel plasma delle code delle comete c'é invece per lo piú ossido di carbonio ionizzato (CO^+) il quale ha una sezione d'urto di diffusione dieci volte maggiore di quella dell'idrogeno. Da calcoli teorici è risultato però che sarebbero necessarie parecchie tonnellate di ossido di carbonio per generare una nube artificiale di CO visibile da grande distanza. Nel 1961 non si poteva disporre in Europa di razzi cosí potenti da consentire di portare tonnellate di carico utile nello spazio, e si è dovuto, perciò, cercare un'altra sostanza che soddisfacesse a una serie di proprietà chimico-fisiche.

Nello spazio la sorgente di energia piú adatta (e per di piú gratuita) per la ionizzazione di un atomo, è il Sole. La sua radiazione ultravioletta è sufficiente per strappare uno o piú elettroni a un atomo che abbia una sezione di fotoionizzazione abbastanza elevata. La luce solare serve inoltre a illuminare la nube, una volta che questa sia stata generata, e a renderla visibile dalla Terra.

L'elemento chimico deve quindi soddisfare alle seguenti proprietà: 1) gli ioni e gli atomi neutri devono avere linee di risonanza nel visibile (devono, cioè, assorbire energia solare ed emetterla a una lunghezza d'onda visibile), in modo che la luce da essi diffusa possa attraversare la « finestra ottica » dell'atmosfera terrestre e giungere agli strumenti di osservazione a Terra; 2) l'elemento deve essere ionizzato in un intervallo di tempo molto breve (la probabilità di ionizzazione deve cioè essere molto alta), poiché in caso contrario la nube neutra diffonde e svanisce prima di essere ionizzata; 3) l'elemento chimico deve essere facilmente evaporabile: la temperatura necessaria per l'evaporazione deve essere pertanto la piú bassa possibile.

Selezionando gli elementi del sistema periodico si trova che soltanto pochi di essi potrebbero soddisfare alle condizioni richieste: il bario (Ba), lo stronzio (Sr), l'europio (Eu) e l'itterbio (Yb). Il comportamento di questi elementi, in particolare del bario e dello stronzio, è stato studiato a fondo in laboratorio; dopo una lunga serie di esperimenti nella grande camera a vuoto per la simulazione spaziale (*si veda l'illustrazione a pagina 56*), si è giunti a una miscela ideale di bario e di ossido di rame con l'aggiunta dello 0,4 % di NaN_3. L'evaporazione del bario si raggiunge in seguito alla seguente reazione chimica:

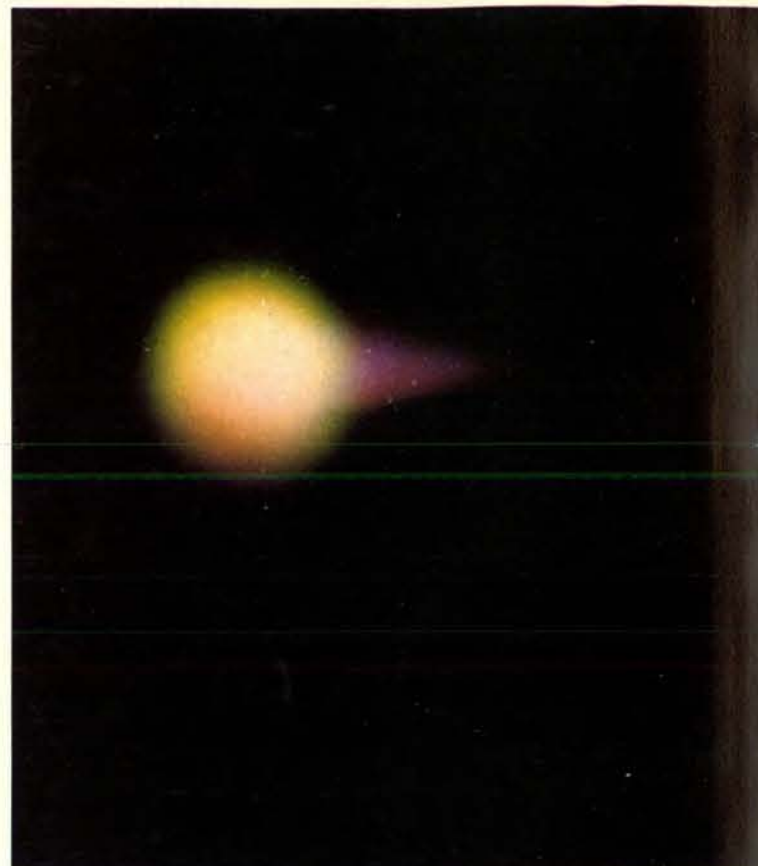
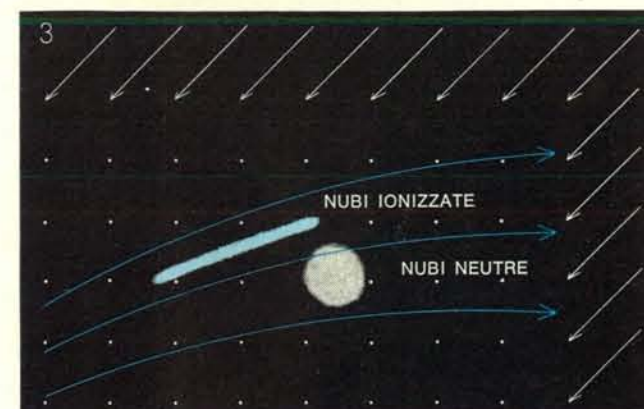
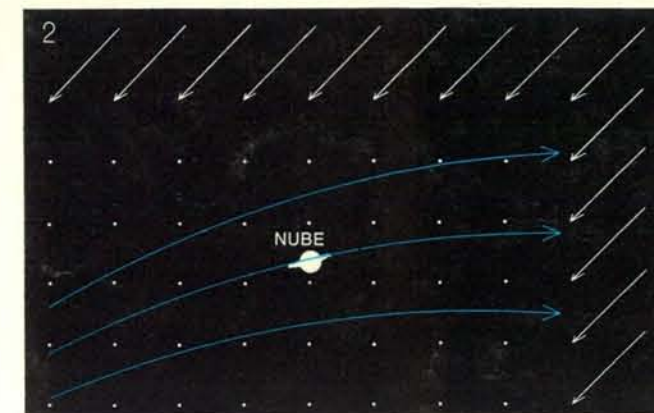
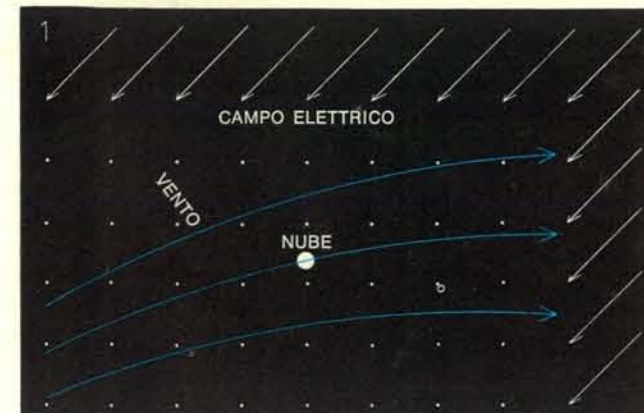
$(n + 1)\ Ba + CuO \rightarrow BaO + Cu + n Ba$ (vapore)

in cui una parte del bario viene bruciata mentre l'ossido di rame funge da ossidante e l'energia che deriva dalla combustione viene usata per l'evaporazione del bario rimanente. La miscela esplosiva viene compressa a strati, per consentire una distribuzione omogenea, entro una capsula dotata di spoletta: una volta iniziata la reazione, i prodotti della combustione fuoriescono attraverso un ugello di Laval nel vuoto a velocità supersonica e a una temperatura di 2500 °K. In questo modo circa il 20 % del bario viene evaporato ed esposto alla radiazione solare. Per ottenere una nube di plasma visibile per piú di un'ora sono necessari appena dai dieci ai cento grammi di ioni.

Molto probabilmente la ionizzazione degli atomi di bario avviene in due fasi: una fase iniziale di circa dieci secondi, che è particolarmente efficace e che dà origine a nubi ioniche molto concentrate, seguita da una fase piú lunga in un intervallo di tempo di cento o piú secondi. Il processo di ionizzazione, che è alquanto complesso, è stato chiarito da Leo Haser nella seguente maniera: 1) assorbimento di risonanza e susseguente occupazione di un livello metastabile in seguito a transizioni spontanee; 2) fotoionizzazione in seguito ad assorbimento nello stato metastabile. (Quando si parla di stato metastabile dell'atomo, si intende un particolare stato eccitato nel quale l'atomo, se non viene disturbato dall'esterno, può rimanere per un periodo di tempo relativamente lungo.) La ionizzazione della nube, inizialmente neutra, oltre che spettroscopicamente, può essere notata anche a occhio nudo, giacché cambiano sia il colore sia la forma.

La nube neutra di bario emette linee spettrali verdi, gialle e rosse; di questa la verde è la piú intensa (linea di risonanza a 5535 Å) e pertanto piú a lungo osservabile, allorché la nube neutra, in seguito a ionizzazione diventa sempre piú debole. Il bario ionizzato ha linee spettrali violette, blu e rosse e produce una nube purpurea (linee di risonanza a 4554 e 4934 Å); una piccola percentuale di stronzio (fino allo 0,5 %) è sufficiente a creare una nube neutra di colore blu (linea di risonanza a 4607 Å), la quale non viene ionizzata dalla radiazione solare e rimane a lungo visibile. Si potrebbe definire un caso fortunato la presenza minima dello stronzio; infatti in questo modo si ha a disposizione una nube artificiale neutra il cui comportamento è interessante per studi di vario genere nell'alta atmosfera. Anche gruppi di altri paesi, oltre che tedeschi, usano ora nubi neutre a tale scopo ed è proprio lo stronzio uno degli elementi preferiti.

Poiché la nube neutra interagisce soltanto con l'atmosfera, si possono studiare molti parametri fisici di quest'ultima osservando il comportamento della nube. Dopo l'evaporazione, la nube neutra comincia a espandersi in tutte le direzioni con una velocità di circa 1200 metri al secondo; dopo poco tempo si osserva la formazione di un involucro sferico dovuto alla presenza di particelle solide che trasferiscono la loro energia alle particelle gassose, anche dopo l'evaporazione. Queste particelle solide, che hanno un diametro variabile fra 0,1 e 50 micron, fuoriescono dall'ugello a una velocità di cento metri al secondo.

L'espansione viene frenata dall'urto fra gli atomi neutri e gli atomi e le molecole dell'atmosfera circostante. Poiché la probabilità di collisione dipende dalla densità dell'atmosfera, il diametro della nube, al momento in cui si è raggiunto l'equilibrio con l'atmosfera circostante, può essere usato per determinare tale densità. Quando l'espansione della nube neutra cessa, gli atomi che si trovano in equilibrio con l'atmosfera cominciano a diffondere verso l'esterno e il processo di diffusione dipende dalla densità e dalla temperatura dell'atmosfera. Determinando il cosiddetto « coefficiente di diffusione », otteniamo facilmente la temperatura dell'atmosfera giacché la densità ci è nota dal processo iniziale di espansione. Un'altra grandezza che si può misurare per mezzo delle nubi neutre, è la velocità dei venti dell'alta atmosfera; ad altezze variabili fra i 125 e 250 chilometri si hanno venti

Nella notte polare la nube di bario ionizzato diffonde la luce solare in modo da essere osservabile anche a occhio nudo; infatti essa assume il tipico colore purpureo del bario ionizzato (Kiruna, Lapponia).

Quattro fasi dello sviluppo di una nube di bario corrispondenti agli schemi della pagina a fronte. In alto a sinistra la nube subito dopo l'evaporazione della miscela. In alto a destra, dopo qualche secondo la radiazione ultravioletta del Sole comincia a ionizzare gli atomi di bario; il colore cambia perché gli ioni di bario, rispetto agli atomi, hanno le linee spettrali piú intense nel violetto, blu e rosso. La parte ionizzata si allunga in direzione del campo magnetico. In basso: a sinistra, la nube dopo 150 secondi dall'evaporazione; a destra la fase di deriva (*drift*). Dopo circa 5 minuti dall'evaporazione le due nubi si allontanano l'una dall'altra, e quella ionizzata di bario si muove sotto l'azione di un campo magnetico e di uno elettrico.

Schemi del moto e dello sviluppo di nubi neutre e ionizzate, corrispondenti alle figure a fronte: *a*) la miscela di bario e ossido di rame è fatta esplodere: si genera una nube di atomi neutri di bario; *b*) inizia l'ionizzazione da parte della radiazione UV del Sole; *c*) la nube di bario ionizzata si stacca da quella neutra di stronzio e si muove sotto l'azione del campo elettrico E e del campo magnetico B; *d*) fase di deriva. Il campo elettrico E va pensato perpendicolare al disegno.

con corrispondenti velocità fra i 15 e 130 metri al secondo. Questi venti non soffiano costantemente a tutte le latitudini come è dimostrato dalle deformazioni subite dalle nubi.

Senza dilungarci ulteriormente sullo studio delle nubi neutre, argomento oramai noto da tempo, torneremo a parlare delle ben piú importanti nubi di plasma.

Come è stato ricordato piú sopra, in seguito all'ionizzazione della nube neutra da parte dell'intensa radiazione ultravioletta del Sole, cambia non solo il colore, ma anche la forma della nube di bario.

Gli ioni di bario e gli elettroni liberi iniziano subito la loro danza elicoidale intorno alle linee del campo magnetico, secondo le leggi fisiche illustrate precedentemente. Il diametro della nube di plasma cresce soltanto in direzione delle linee del campo magnetico e, conseguentemente, la nube passa da una forma circolare a una ellittica e infine sigariforme; può essere quindi facilmente distinta da quella neutra come mostrano le illustrazioni in queste due pagine.

In questo modo vengono rese visibili le linee del campo magnetico terrestre allo stesso modo che la polvere di ferro, sparsa su un foglio di carta, rende visibili le linee del campo magnetico di una calamita. La nube di plasma interagisce anche con i campi elettrici presenti nello spazio extraterrestre e compie quindi un movimento di deriva perpendicolare ai due campi. Proprio da questo moto di deriva e dal valore del campo magnetico (che è noto e che si aggira normalmente intorno agli 0,5 gauss) si è potuto per la prima volta determinare il valore dei campi elettrici a varie latitudini.

Misurare un campo elettrico nello spazio, per mezzo di un satellite artificiale o una sonda spaziale, non era stato possibile per due motivi: i campi elettrici sono debolissimi (anche minori di 1 volt al chilometro), il veicolo spaziale influenza, con la sua presenza, in modo determinante i campi elettrici che esso incontra; infatti muovendosi nel plasma circostante, che è magnetizzato, vi induce un campo elettrico che è, per lo piú, maggiore del campo che si vuole misurare.

Magnetosfera e magnetopausa

La materia interplanetaria non rappresenta, fino a distanze di 50-100 unità astronomiche dal Sole, altro che la continuazione della corona solare; questa parte esterna della corona si trova in uno stato di continua espansione, il cui carattere si può chiarire su base prettamente idrodinamica e ha una grande somiglianza con l'accelerazione di un gas surriscaldato attraverso un ugello di Laval.

Sebbene il campo magnetico terrestre non abbia un influsso importante dal punto di vista energetico, la sua presenza è decisiva per l'interazione con il vento solare, che può essere paragonato all'acqua che investe la prua della nave e da questa viene deviata: da tale similitudine è nata la definizione d'onda di prua (*bow wave*) per indicare l'onda d'urto che viene creata dal vento solare a contatto con il campo magnetico terrestre (*si veda l'illustrazione a pagina 66*). L'onda d'urto comprime le linee del campo magnetico terrestre fino a una distanza di circa 60 000 chilometri nella parte ri-

volta al Sole, e le « soffia via » fino ad una distanza di circa 600 000 chilometri, nella parte opposta al Sole; questa regione dello spazio extraterrestre, in cui il movimento della materia è fortemente influenzato dalla presenza del campo magnetico terrestre, è chiamata magnetosfera e il suo confine magnetopausa. La magnetosfera somiglia a una pianta che ha le sue radici nella Terra, trae energia in parte dalla Terra, in parte dal Sole e ha una forma definita, sebbene la sua materia si trovi in continuo movimento. Essa si muove nel vento solare e cambia la sua forma a seconda dell'intensità dell'attività solare; non rimane immutata nel tempo, bensí sembra dissolversi e rinascere continuamente.

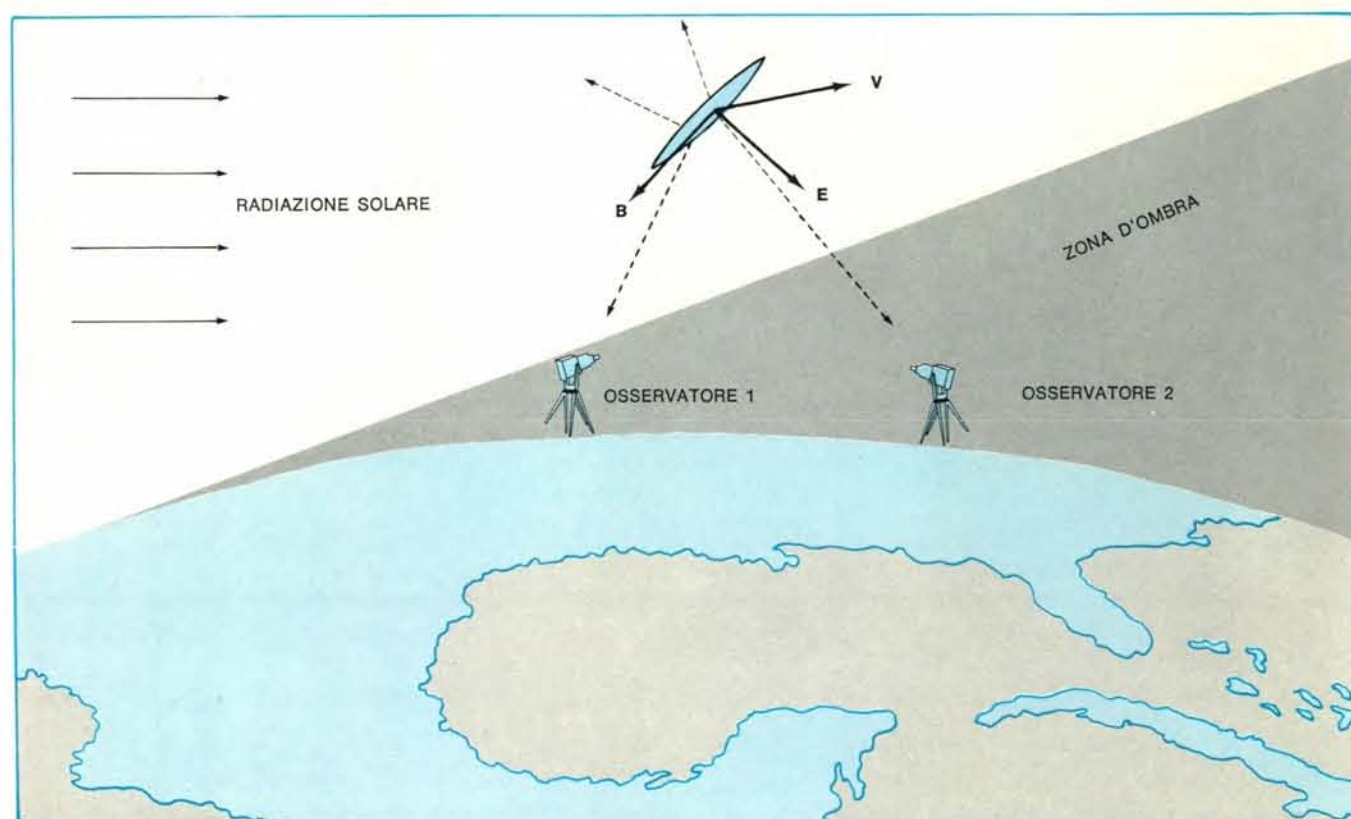
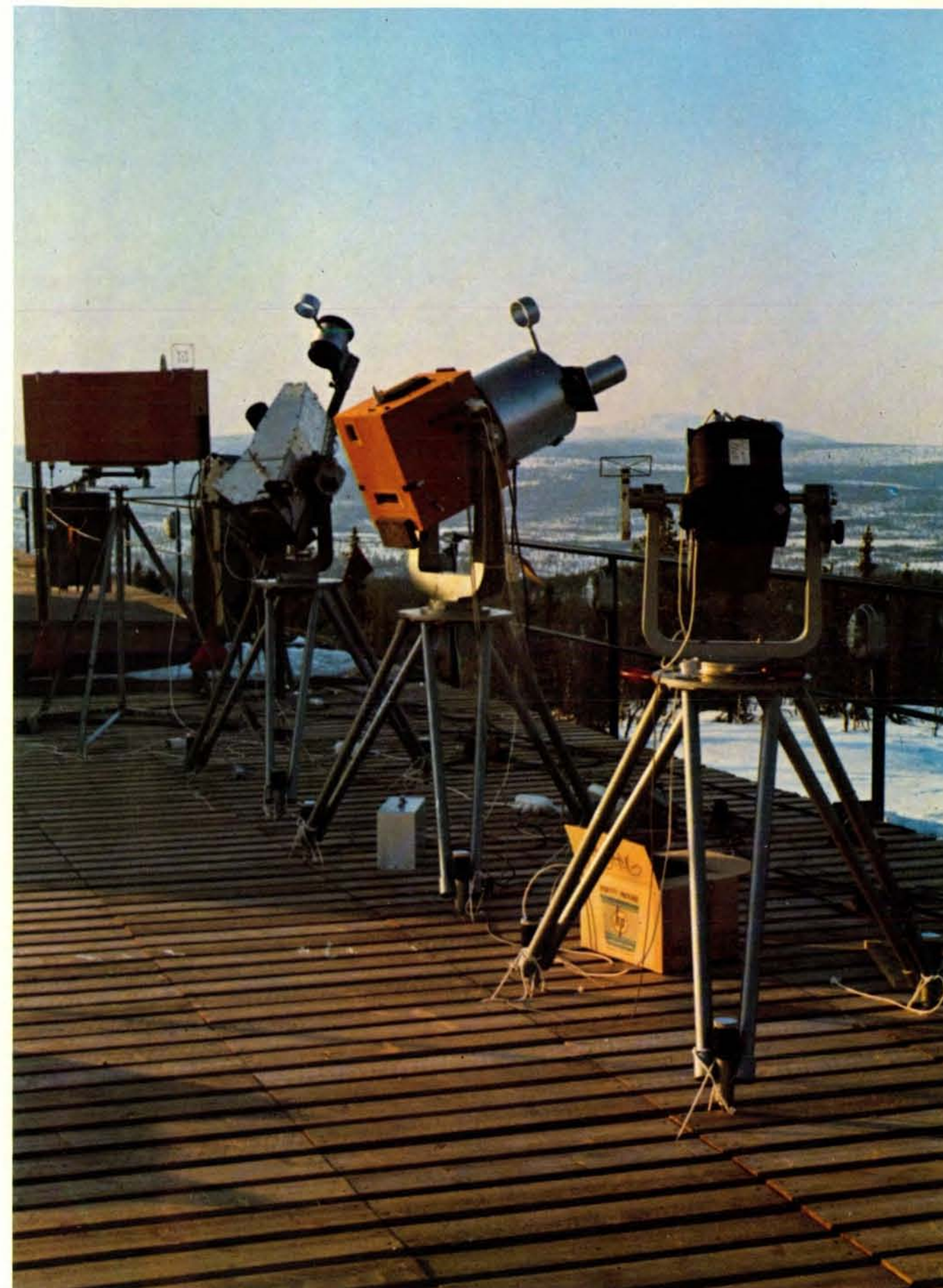
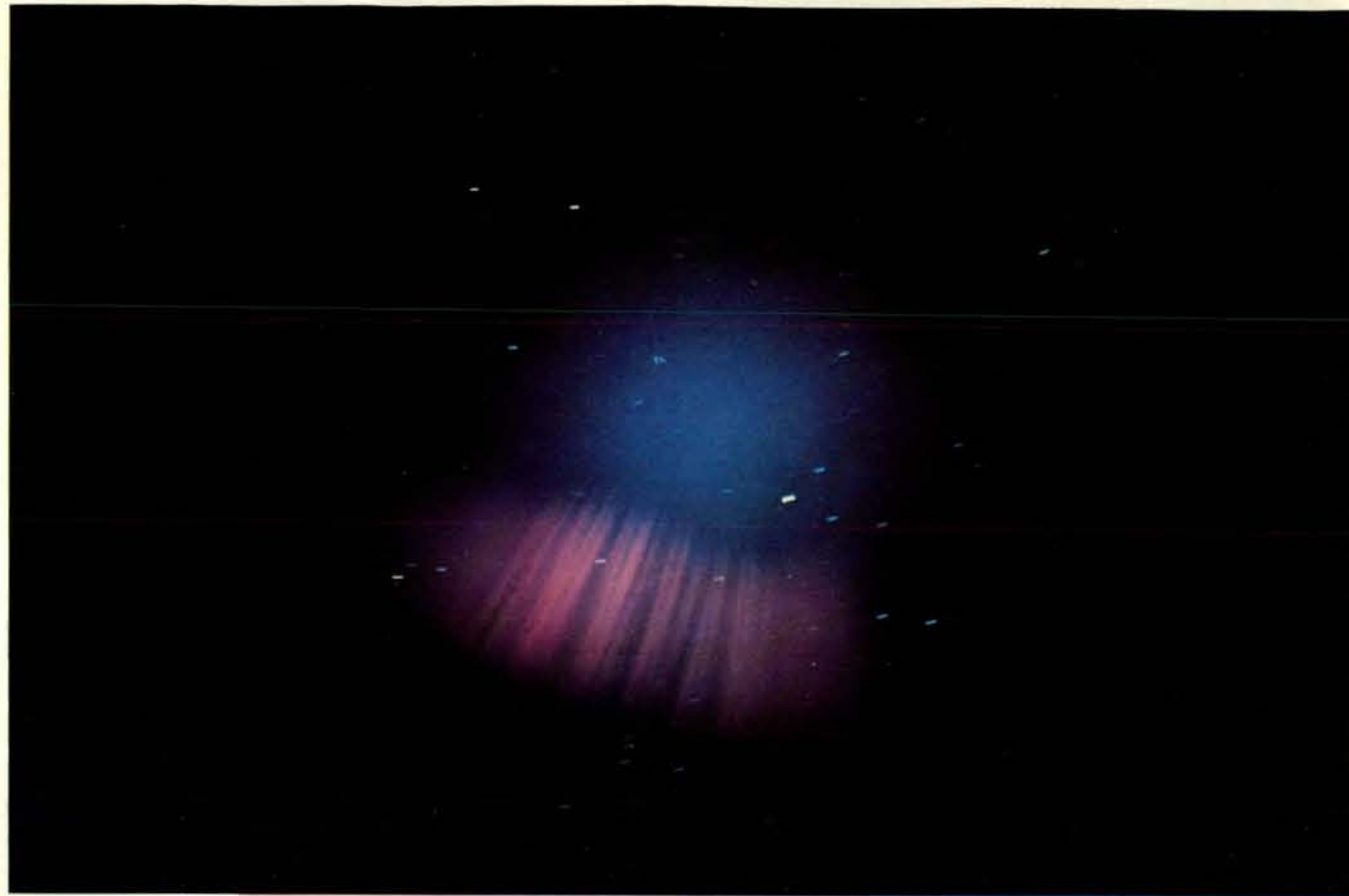
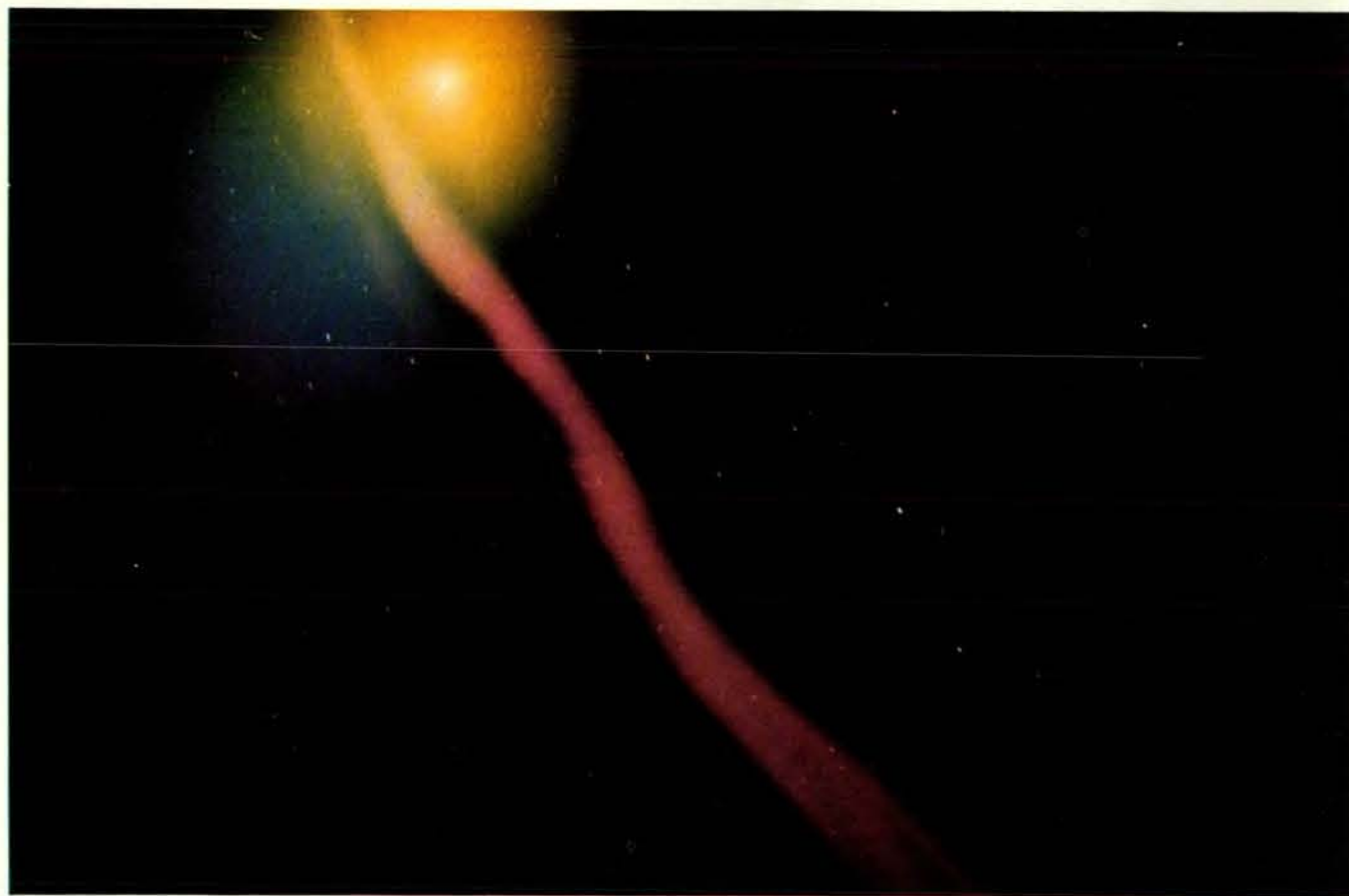
La coda della magnetosfera è simile alla coda di una cometa in quanto ambedue, come una banderuola al vento, vengono spinte in direzione opposta al Sole; però, a differenza della coda di una cometa, la coda magnetica non è otticamente visibile, giacché la materia di cui è formata consta soltanto di idrogeno ionizzato, ma non luminoso; al contrario la luminosità delle code delle comete è dovuta alla diffusione della luce solare da parte di ioni molecolari come CO^+ che hanno una sezione d'urto di diffusione dieci volte maggiore di quella dell'idrogeno.

I movimenti convettivi del plasma nella magnetosfera, causati dal vento solare, rappresentano il fattore determinante per la generazione di campi elettrici nella magnetosfera stessa. Una serie di interessanti fenomeni osservati nelle particelle che costituiscono le fasce di van Allen e soprattutto nelle particelle presenti nelle aurore boreali, sono condizionati da questi campi elettrici; essi si propagano lungo le linee del campo magnetico fino alla bassa ionosfera, ove fungono da forza motrice per le correnti elettriche che qui vengono registrate.

I campi elettrici nella ionosfera

La determinazione dei campi elettrici nella ionosfera (regione che si estende fra i 70 e i 1000 chilometri circa di altezza e che rappresenta la zona di transizione fra l'atmosfera e la magnetosfera) ci può fornire anche utili informazioni sul comportamento e sulla intensità di campi elettrici a grande distanza dalla Terra e sull'interazione del vento solare con la magnetosfera. Questo è stato anche uno dei motivi principali per la produzione di nubi di plasma in tale regione.

Teoricamente i campi elettrici si determinano dalla variazione, osservata da Terra, del campo magnetico terrestre. La premessa per queste misure, è la conoscenza del rapporto fra intensità di corrente e campo elettrico per tutti gli strati sovrastanti la superficie terrestre, vale a dire la conoscenza della conducibilità elettrica. Questi rapporti sono, proprio nella ionosfera, particolarmente complessi per il fatto che le particelle neutre dell'atmosfera influiscono sulla conducibilità elettrica (grandezza fisica che in questo caso non è uno scalare, bensí un tensore). Negli strati della ionosfera a partire dai duecento chilometri di altezza, la frequenza di rotazione degli ioni e degli elettroni è elevata rispetto alla frequenza d'urto delle particelle cariche con le particelle neutre e in questo caso il plasma è soggetto al movimento di deriva precedentemente descritto.

Al diminuire dell'altezza, la frequenza di rotazione è uguale alla frequenza d'urto; ciò significa che il campo magnetico non è piú cosí efficace e il moto di deriva trasversale decresce. Di conseguenza gli ioni e gli elettroni cominciano a muoversi in direzione del campo elettrico, come farebbero in as-

Triangolazione della nube di plasma di cui si osserva l'espansione con macchine fotografiche, telecamere e spettrografi da due stazioni di osservazione, a circa 100 km l'una dall'altra. Gli osservatori devono essere in zona d'ombra per vedere la nube, illuminata dal Sole, sul fondo scuro del cielo. Gli ioni di bario diffondono la luce solare a ben determinate lunghezze d'onda assumendone il colore relativo: *B* indica la direzione del campo magnetico, *E* di quello elettrico e *V* quella del moto di deriva.

Stazione di osservazione a Kiruna (Svezia) con spettrografi e macchine fotografiche costruite per l'osservazione delle nubi artificiali. A volte si aspettano settimane per eseguire gli esperimenti nello spazio essendo necessaria una visibilità perfetta.

Esperimento a Fort Churchill (Canada) con due esplosioni alla distanza di tre minuti (*sopra*). Quando si forma la seconda nube (gialla), la prima è già decomposta in una neutra di stronzio (blu) e in una ionizzata di bario (purpurea), della lunghezza di circa cento chilometri, le cui striature (*sotto*), allineate con le linee del campo magnetico, sono dovute a una distribuzione non omogenea del plasma nella ionosfera e nella magnetosfera: tali striature sono larghe da uno a dieci chilometri.

senza di un campo magnetico. Poiché, però, queste particelle hanno cariche opposte, saranno anche opposte le componenti delle loro velocità: si avrà quindi una corrente elettrica, detta corrente di Pedersen (dal nome del fisico danese che introdusse questo concetto).

Tale corrente, causata praticamente dagli ioni, scorre soprattutto ad altezze superiori ai 110 chilometri. Ad altezze inferiori il rapporto tra frequenza di rotazione e frequenza d'urto diminuisce con la conseguenza che tutti gli effetti magnetici scompaiono e le particelle cariche seguono il moto delle particelle neutre. Per gli elettroni questo rapporto è molto maggiore di quanto non lo sia per gli ioni; c'è pertanto una regione della ionosfera fra gli 80 e i 110 chilometri di altezza, ove gli ioni si accompagnano alle particelle neutre, seguendo lo stesso loro movimento, mentre gli elettroni continuano indisturbati il moto di deriva. La corrente elettrica che cosí ha origine corrisponde alla ben nota corrente di Hall nei metalli.

In base agli effetti descritti, è chiaro che la conoscenza del rapporto fra corrente elettrica e campo elettrico nella ionosfera è piuttosto difficile, poiché alla normale conducibilità elettrica si aggiungono le cosiddette conducibilità di Hall e di Pedersen. Queste ultime possono causare un cambiamento anche di 90° nella direzione della corrente elettrica rispetto a quella del campo elettrico.

Oltre a questi campi elettrici di origine extraterrestre, ci sono nella ionosfera anche campi elettrici di origine terrestre, generati dal movimento dell'atmosfera attraverso il campo magnetico e dalla rotazione della Terra.

Questo effetto, che fu studiato già nel secolo scorso da B. Stewart, ma che non era stato ancora confermato, viene chiamato « effetto dinamo ». Infatti l'atmosfera neutra, che si muove attraverso il campo magnetico terrestre, si può paragonare alla spira di una dinamo che, ruotando, genera una corrente elettrica. Analizzando ora i campi elettrici misurati per mezzo delle nubi di plasma, si verifica la presenza di correnti elettriche e l'esattezza di una teoria formulata nel lontano 1882.

A latitudini medie (fra il Sahara e la Sardegna) e all'ora del crepuscolo – poiché tali correnti variano con l'ora solare – si è potuto stabilire che il campo elettrico ha direzione opposta alla corrente elettrica; la direzione della corrente elettrica si può determinare dall'osservazione del campo magnetico alla superficie terrestre.

A queste latitudini sono stati eseguiti esperimenti sopra Hammaguir (Sahara), Perdasdefogu (Sardegna) e sopra le isole Wallops (USA). I razzi usati sono stati dei Centaure e dei Dragon francesi e lo Skylark inglese per esperimenti compiuti fra i 135 e i 230 chilometri di altezza, mentre per esperimenti a 1000 e 2000 chilometri sono stati usati rispettivamente lo Javelin americano e il Rubis francese.

L'intensità dei campi elettrici misurati a queste altitudini, varia fra un minimo di 1 volt al chilometro e un massimo di 5 volt al chilometro; tale massimo è stato misurato a 220 chilometri d'altezza sul cielo della Sardegna. Questi campi elettrici sono molto deboli in paragone ai campi elettrici conosciuti sulla Terra (per esempio, quelli generati durante un forte temporale, raggiungono un'intensità dell'ordine di un miliardo di volt al chilometro).

A partire dal 1966, oltre che nelle regioni di media latitudine, sono stati effettuati esperimenti con nubi di bario a Kiruna (Lapponia) e a Fort Churchill (Canada), (*si vedano le illustrazioni alle pagine 59, 63, 64*), vale a dire nella zona polare nordica, ove lo studio delle correnti elettriche e dei campi elettrici è di particolare importanza, data la vicinanza dei poli magnetici. Qui hanno origine le linee del campo magnetico della coda, la quale, come abbiamo visto, è fortemente influenzata dall'interazione con il vento solare.

I movimenti del plasma nella coda e, di conseguenza, i movimenti delle linee del campo magnetico, a causa dell'alta conducibilità elettrica nella magnetosfera, generano campi elettrici; questi possono accelerare elettroni e conferire loro un'energia dell'ordine dei keV, sufficiente per penetrare nell'atmosfera terrestre. Gli elettroni urtando con le molecole e gli atomi neutri, li eccitano. Dopo un intervallo di tempo piú o meno breve, le molecole e gli atomi eccitati irradiano quelle particolari lunghezze d'onda, cioè quei particolari colori, che rendono tanto spettacolari le aurore boreali.

L'aurora boreale e le nubi di plasma

L'aurora boreale, uno dei piú fantastici fenomeni ottici che la natura può offrire, è stata per secoli un problema oscuro, neppure oggi completamente risolto.

È certo però che le forti correnti elettriche, generate secondo il meccanismo precedentemente descritto, e la presenza di particelle cariche giocano una funzione determinante nell'apparizione delle aurore boreali.

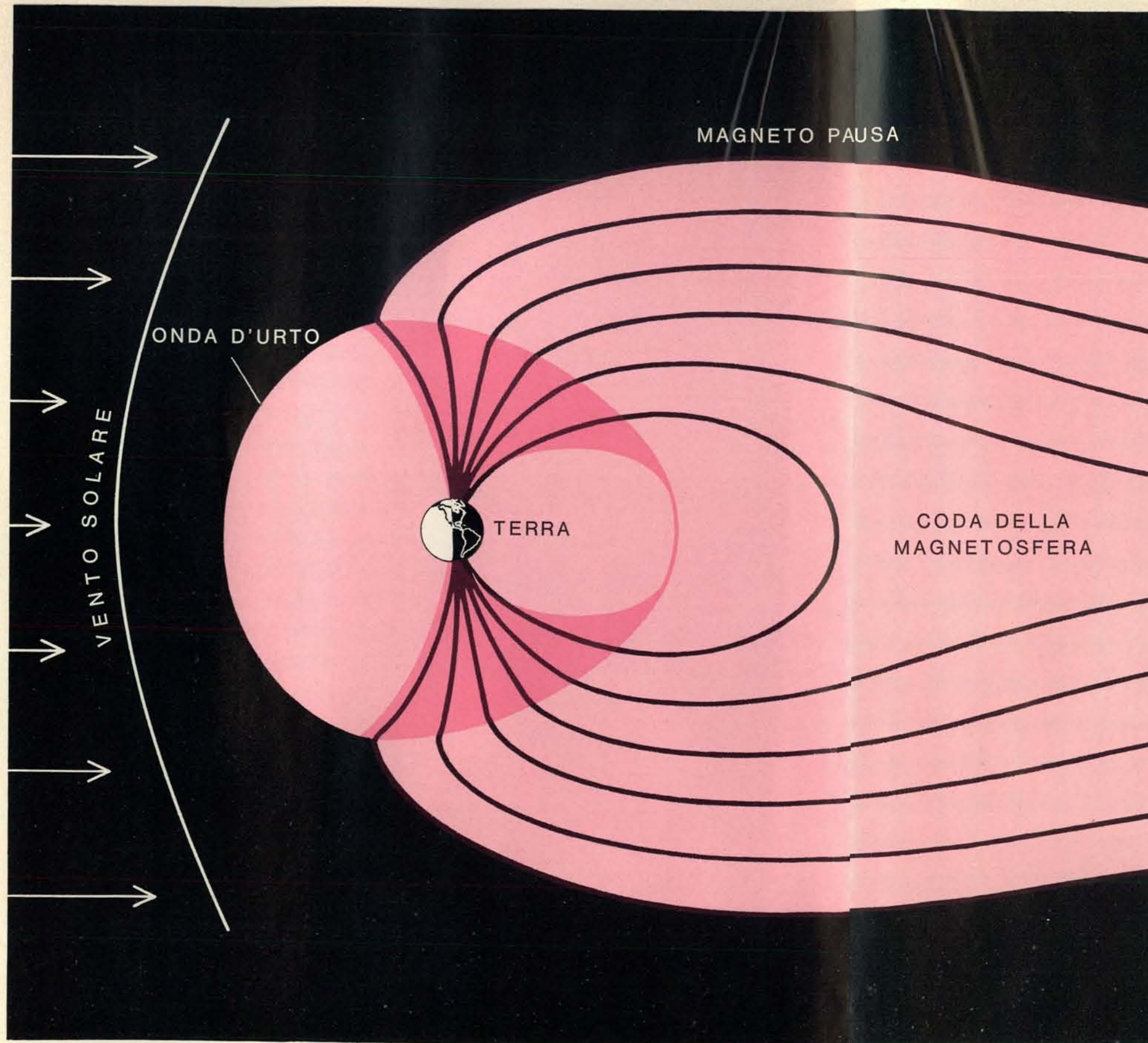
Producendo nubi di plasma artificiale in tale regione si potranno pertanto osservare dei moti di deriva causati prevalentemente dall'interazione indiretta della nube ionizzata con il vento solare.

A Kiruna e a Fort Churchill sono stati eseguiti esperimenti fra i 200 ed i 380 chilometri d'altezza, servendosi di razzi canadesi di tipo Black Brant e americani di tipo Nike-Apache e Nike-Tomahawk.

Per poter determinare l'altezza, l'espansione e la velocità di deriva della nube, e di conseguenza i campi elettrici, è necessario eseguire una triangolazione della nube da due stazioni distanti tra loro per lo piú cento chilometri. Queste stazioni, messe a disposizione dall'ESRO o allestite direttamente dall'Istituto Max-Planck, e dotate di spettrografi, camere televisive e camere fotografiche speciali, devono trovarsi in zona d'ombra al momento dell'evaporazione della nube la quale viene irradiata dal Sole già tramontato (*si veda l'illustrazione a pagina 62*). La piú grande preoccupazione durante le nostre spedizioni nelle regioni polari è il fattore meteorologico, giacché le condizioni di visibilità devono essere perfette per eseguire tali esperimenti; è successo talvolta, con nostro grande disappunto, di dover tornare indietro a mani vuote, dopo aver aspettato settimane intere la scomparsa delle nubi naturali per poter generare quelle artificiali!

Oltre alle capsule contenenti la miscela di bario e ossido di rame sono stati aggiunti al carico utile dei razzi strumenti per misurare la densità elettronica, in modo da poter ottenere informazioni sulla conducibilità ionosferica; ciò per consentire un paragone fra le perturbazioni magnetiche, osservate da Terra, e quelle dovute alle correnti provocate dai campi elettrici misurati. Inoltre può, cosí, essere apprezzata la perturbazione che la nube stessa genera nella conducibilità originaria.

Dai risultati ottenuti con gli esperimenti di Kiruna e Fort Churchill si è potuto verificare che le perturbazioni del campo magnetico, calcolate teoricamente in base alle correnti elettriche, coincidono con le perturbazioni osservate direttamente.

La conducibilità massima viene registrata a un'altezza di cento chilometri, ove la corrente di Hall è predominante. Poiché tale corrente è generata da elettroni che si muovono in senso opposto, la nube artificiale viaggerà verso est ogni volta che il magnetometro, a Terra, registrerà una corrente di Hall che scorre nella ionosfera in direzione ovest e viceversa.

Magnetosfera terrestre: il vento solare non può penetrare nell'atmosfera terrestre a causa dell'onda d'urto che si genera in seguito alla sua interazione con il campo magnetico terrestre. Le linee del campo magnetico sono molto influenzate da quest'interazione e, mentre quelle esposte al Sole vengono compresse fino a una distanza di circa 60 000 chilometri dalla Terra, quelle opposte al Sole vengono soffiate via formando una coda magnetica di circa 600 000 chilometri di lunghezza.

La velocità della nube di bario varia sia in grandezza sia in direzione e ciò mostra che il campo elettrico è variabile. Sono stati misurati valori dell'intensità del campo elettrico che oscillano fra 1 e 130 volt per chilometro, con punte massime cioè piú di 20 volte superiori a quelle registrate nelle zone di media latitudine.

Particolarmente interessante è la configurazione delle nubi di plasma in vicinanza degli archi delle aurore boreali; esse si muovono quasi parallele a questi archi, assumendo una forma longilinea che raggiunge anche duecento chilometri di lunghezza (*si veda l'illustrazione a pagina 64, in basso*). Eventuali movimenti degli archi, perpendicolari alla loro direzione longitudinale sono imitati dalle nubi di plasma.

Rispetto all'arco dell'aurora in quiete, il campo elettrico è perpendicolare all'arco stesso. Questo risultato è molto significativo, perché ci dimostra che la materia nella magnetosfera si muove parallela agli archi delle aurore boreali.

Anche la corrente di Hall scorre lungo tali archi, che devono essere considerati come zone di alta conducibilità.

Non è però ancora stato chiarito in che modo vengano a richiudersi le correnti ionosferiche. Da un lato è certo che ciò si verifica in parte entro la magnetosfera.

Nuove misure di campi elettrici, con la contemporanea misura di altri parametri e con una fitta rete di osservazioni da Terra, serviranno in futuro a far piú luce sui fenomeni ancora oscuri che si manifestano con le aurore boreali.

Un particolare fenomeno si osserva nelle nubi ionizzate a qualche minuto dall'evaporazione: delle striature, di larghezza variabile, appaiono in direzione delle linee del campo magnetico. Inomogeneità entro il plasma sembrano essere la causa di questo fenomeno, che si riscontra anche nelle code delle comete; esso è molto complesso e i teorici del plasma nel nostro istituto ne hanno fatto oggetto di intenso studio. La larghezza delle striature oscilla fra 1 e 10 chilometri e non è costante nel tempo (*si veda l'illustrazione a pagina 64, in alto*).

L'uso delle cariche cave per generare nubi di plasma

Recentemente è stato sperimentato dal nostro gruppo, sotto la guida dello scienziato tedesco Karl-Wolfgang Michel, un nuovo metodo per generare nubi di plasma nello spazio: quello delle cosiddette « cariche cave ». Una carica cava (*si veda l'illustrazione a pagina 68*) consiste di un involucro cilindrico di TNT e RDX (materiali altamente esplosivi) in cui è scavata una cavità conica rivestita da uno strato di bario.

Da tempo si conosceva la possibilità di accelerare a grandi velocità una parte del rivestimento metallico delle cariche cave e, durante la seconda guerra mondiale, questo procedimento venne usato per la distruzione di mezzi corazzati.

Il nostro problema consisteva nel generare fasci di ioni veloci che, « sparati » a un'altezza di circa duecento chilometri, potessero raggiungere i quattrocento chilometri o piú e ivi seguire a grande velocità le linee del campo magnetico. Per fare questo bisognava riscaldare il bario a una temperatura superiore ai 2000 °C (come nel caso delle miscele classiche), per farlo evaporare, e accelerarlo a una velocità di dieci chilometri per secondo. L'energia necessaria a tale scopo può essere fornita da una carica cava. Un'onda piana di detonazione, causata dall'esplosione della miscela (80 % RDX e 20 % TNT), trasmette un'onda d'urto al bario compresso e fornisce una temperatura di 6500 °C a una pressione di ben 300 000 atmosfere. Le onde d'urto trasversali di una carica cava classica non sono però abbastanza efficaci per ottenere il necessario quantitativo di vapore di bario alla velocità di dieci chilometri per secondo. Dopo un anno di esperimenti nella grande camera a vuoto si è realizzata una carica cava ideale che permette, tramite la generazione di una seconda onda d'urto, di ottenere le condizioni richieste.

Questo metodo, che oltrettutto non necessita di grandi razzi ed è pertanto economico, apre nuove interessanti prospettive per lo studio della visualizzazione delle linee del campo magnetico e dell'interazione fra ioni veloci e molecole dell'alta atmosfera. Si pensa anche di « sparare », con questo metodo, un fascio di ioni veloci dall'emisfero nord a quello sud lungo una delle linee chiuse del campo magnetico terrestre.

L'esperimento HEOS e le prospettive per il futuro

Nel marzo del 1969 abbiamo avuto la prima grande occasione per sperimentare il metodo delle nubi artificiali di plasma a grande distanza dalla Terra.

Il satellite HEOS A, costruito per

conto dell'ESRO (Organizzazione europea per la ricerca spaziale), aveva a bordo, oltre a strumenti di grande sensibilità progettati da vari gruppi europei con differenti scopi di ricerca, una capsula di 2400 grammi di miscela esplosiva. Il satellite, che aveva un'orbita prettamente eccentrica, con un perigeo di 424 e un apogeo di 223 428 chilometri di media – infatti l'orbita viene alterata a causa dell'azione della Luna – ha sganciato la capsula a una distanza di circa 74 000 chilometri dalla Terra, in una zona della magnetosfera esterna prescelta per l'osservazione da Terra. Dopo tre ore circa a una distanza di 30 chilometri dal satellite (onde evitare un'eventuale contaminazione degli apparecchi di bordo), è stata generata la nube artificiale.

Essa è stata seguita per 27 minuti dalle stazioni di osservazione appositamente allestite a La Serena (Cile) e sul Kitt Peak (Arizona), e dotate di telescopi Super-Schmidt. Inoltre la nube dell'HEOS è stata fotografata da Monte Palomar con il telescopio Schmidt da 48 pollici, da Fairbanks (Alaska) con camere televisive, da Mt. Hopkins (Arizona), Arequipa (Perù) e Natal (Brasile) con camere Baker-Nunn.

L'ampiezza dell'angolo solido entro la carica cava (*a*) è determinante per la velocità e la temperatura del vapore di bario. Quando la miscela esplosiva (*b*) detona, si genera dopo pochi microsecondi un getto di bario surriscaldato, che si disgrega dopo circa 20 μs (*c*) sotto l'azione dell'onda d'urto e diventa fluido. Per l'alta temperatura alla velocità di 10 km/s, evapora e forma un getto di atomi veloci di bario che, poi, sia per urto sia a causa della radiazione solare, si trasforma in un fascio di ioni veloci.

La densità della nube, durante il periodo di osservazione, si è aggirata su valori che vanno da $5 \cdot 10^4$ particelle per centimetro cubo fino a 10^3 particelle al centimetro cubo, mentre la densità del plasma magnetosferico oscillava fra 1 e 10 particelle per centimetro cubo.

Dopo 20 secondi la nube è stata ionizzata dalla radiazione solare e i 100 grammi di ioni di bario prodotti si sono disposti in direzione delle linee del campo magnetico; in tale direzione la nube di plasma si è allungata per circa 5000 chilometri, mentre in direzione perpendicolare al campo magnetico aveva una larghezza approssimativa di 100 chilometri.

Dopo 540 secondi dall'evaporazione la densità entro la nube è circa 10 000 volte maggiore della densità del plasma circostante. Non è pertanto sorprendente il fatto che, fino al termine dell'osservazione, la parte centrale della nube abbia mantenuto immutata la sua velocità iniziale, mentre le zone esterne, influenzate dal movimento del plasma circostante, sono rimaste indietro in direzione ovest. Pertanto in questa direzione la nube appare sfumata.

Si è cosí venuto a formare qualcosa di analogo a una coda di cometa. La direzione della coda della nube ha dato informazioni sulla direzione dei campi elettrici nella magnetosfera.

È stato inoltre osservato che la nube presentava striature parallele alle linee del campo magnetico terrestre; il processo che le genera è però, in questo caso, da paragonarsi solo lontanamente con quello della ionosfera. Nel caso della nube dell'HEOS le striature sono dovute alla cosiddetta *flute instability*, una instabilità che sembra giocare una funzione importante anche nell'astrofisica.

Nel debole campo magnetico di $5 \cdot 10^{-4}$ gauss predomina, nello stato iniziale della nube, per circa 30 secondi, la pressione del plasma di bario in espansione rispetto alla pressione del campo magnetico. Di conseguenza il campo magnetico nell'interno della nube viene indebolito, com'è stato accertato dal magnetometro dell'HEOS, il quale si trovava nella zona marginale della nube.

Ci si potrebbe chiedere come proseguiranno gli esperimenti nei prossimi anni e quale fine si vorrà raggiungere con le nubi artificiali di plasma.

A Garching si vorrebbe da una parte studiare determinati problemi della ionosfera, di cui abbiamo visto la complessità, senza perdere di vista lo scopo iniziale, vale a dire eseguire esperimenti a grande distanza dalla Terra per lo studio del plasma entro la magnetosfera e fuori dalla magnetosfera stessa.

Il secondo esperimento nella magnetosfera avrà luogo nel 1971 per mezzo di un razzo Scout, fornito dalla NASA, che porterà per la prima volta una quantità di ben 13 chilogrammi di miscela esplosiva a 30 000 chilometri d'altezza.

Scopo di questo esperimento è innanzitutto lo studio dei campi elettrici, la struttura del campo magnetico e la microstruttura del moto del plasma; in secondo luogo lo studio del comportamento della nube ionica in un plasma ove le collisioni siano trascurabili. Di particolare interesse per questo secondo scopo sono: l'espansione della nube, la sua accelerazione e le perturbazioni che essa genera nella materia circostante.

Studi sul plasma ove le collisioni sono trascurabili, risultano di particolare importanza poiché sembra tuttora molto difficile riuscire a simulare tali condizioni in laboratorio.

Si spera pertanto, con tale esperimento di fare un passo in avanti, per quello che riguarda particolari problemi concernenti un simile plasma.

Il prossimo passo consisterà poi nell'uscire dalla magnetosfera e generare nello spazio interplanetario una nube di plasma che verrà pertanto esposta direttamente al bombardamento del vento solare e, in assenza del campo magnetico terrestre, si comporterà come una vera coda di cometa.

Dalle osservazioni che si eseguiranno sarà quindi possibile risalire al meccanismo di interazione fra vento solare e code di comete, e ciò comporterà l'uso di razzi di grande portata che possano trasportare una grande quantità di miscela esplosiva nello spazio interplanetario; non è da escludere, dato l'interesse degli americani a tale esperimento, che la NASA metta a disposizione degli scienziati europei in un prossimo futuro un razzo di tipo Saturno.

In tale modo si potrà forse scrivere un'altra importante pagina nella giovane storia della ricerca spaziale.

Come ricordiamo ciò che vediamo

Il fenomeno è in relazione al fatto che ciò che vediamo abbia carattere figurativo (per esempio, scene o fotografie) o linguistico (parole, numeri, ecc.). La memoria linguistica e quella figurativa hanno caratteri diversi

di Ralph Norman Haber

La percezione visiva è connessa col ricordo di ciò che abbiamo visto non meno che con l'atto stesso del vedere. Guardando un quadro, io sono consapevole di vederlo e posso descrivere l'esperienza del vedere. Posso anche ricordare ciò che ho visto quando non ho piú il quadro dinanzi a me. Come funziona questo tipo di memoria percettiva? È un processo che implica una sequenza di diversi stadi, oppure un processo a un solo stadio durante il quale possono svolgersi molti processi parallelamente? Le scene, i visi, i quadri vengono ricordati in modo diverso dal materiale linguistico percepito visivamente come per esempio le parole ed i numeri?

Nel tentativo di rispondere a queste domande i miei colleghi dell'Università di Rochester e io, e ricercatori di altri laboratori, abbiamo studiato il processo della memoria visiva nei soggetti umani. Fra i nostri strumenti vi sono il tachistoscopio (che permette di presentare una serie di immagini in successione rapida), i proiettori di diapositive e gli schermi, strumenti che permettono di seguire i movimenti oculari, strumenti che misurano il tempo necessario per rispondere agli stimoli e materiale figurativo e linguistico di vario tipo. Questi esperimenti incominciano a rivelare varie caratteristiche importanti del processo della memoria visiva; fra i reperti piú significativi vi è la probabile esistenza di due tipi di memoria: uno per il materiale figurativo e l'altro per il materiale linguistico.

La capacità di ricordare materiale visivo può essere addirittura illimitata, come dimostra anche l'esperienza comune: per esempio, quasi tutti hanno fatto l'esperienza di riconoscere un viso che avevano visto solo per brevissimo tempo molti anni prima. (Come vedremo, è significativo il fatto che di solito è piú difficile ricordare il nome della persona in questione.) Che queste esperienze corrispondano alla realtà lo dimostrano anche gli esperimenti: in uno di questi i soggetti riuscirono a riconoscere ben 600 immagini che avevano visto solo per un breve momento. Piú di recente, Lionel G. Standing e io abbiamo dimostrato in un esperimento che è possibile riconoscere un materiale per lo meno quadruplo.

Nel nostro test di capacità mnemonica visiva si mostrarono ai soggetti 2560 diapositive al ritmo di una ogni 10 secondi in sedute che avvenivano in giorni consecutivi. Supponendo che la stanchezza avrebbe potuto avere un qualche effetto sulla prestazione, disponemmo che alcuni volontari seguissero un programma intensivo, consistente nel guardare 1280 immagini al giorno in sedute di quattro ore per due giorni consecutivi; i restanti soggetti guardarono solo 640 immagini al giorno in sedute di due ore per quattro giorni consecutivi.

Un'ora dopo che il soggetto aveva visto l'ultima diapositiva, gli si mostravano 280 coppie di immagini: un membro di ogni coppia era un'immagine della serie già vista dal soggetto, l'altro proveniva da una serie simile, ma che non gli era mai stata mostrata. I soggetti dovevano dire quale delle due immagini avevano già visto in precedenza, e le scelte furono corrette nell'85-95 per cento dei casi. Sorprendentemente, la prestazione dei soggetti che erano stati sottoposti alle sedute piú lunghe e faticose non fu inferiore a quella dei soggetti che avevano seguito un programma meno intenso. In un'altra versione dell'esperimento i punteggi rimasero alti anche quando le figure furono proiettate secondo l'immagine speculare nelle sedute di identificazione, in modo che la parte destra veniva a trovarsi a sinistra e viceversa. Solo quando si chiese ai soggetti se le immagini erano state voltate, si ebbe un certo calo delle risposte corrette (*si veda l'illustrazione a pag. 72*).

Una persona può ricordare quasi ogni immagine vista in passato, ma spesso non riesce a ricordare su richiesta i particolari di una immagine specifica. In un altro esperimento, Mathew H. Erdelyi e io ci proponemmo di appurare che cosa accade a questi particolari sfuggiti: non sono mai stati visti? sono stati visti ma poi dimenticati, oppure sono visti e ricordati, ma in modo che il soggetto non riesce a rievocarli in circostanze normali? Per rispondere a questi quesiti si mostrò per breve tempo ad ogni soggetto un'immagine molto ricca di particolari, poi gli si chiese di rievocare sia a parole sia in un disegno tutti gli elementi dell'immagine che ricordava di aver visto. Quando il soggetto disse che non riusciva a ricordare niente di piú, gli ponemmo domande genericamente stimolanti ma non direttive finché la sua capacità di rievocare ulteriori particolari sembrava proprio esaurita. (Per esempio: « Qui ha disegnato un uomo in piedi: saprebbe descrivere come era vestito? » « Ha lasciato vuoto l'angolo inferiore destro del suo disegno: ricorda cosa c'era nell'immagine in questa zona. »

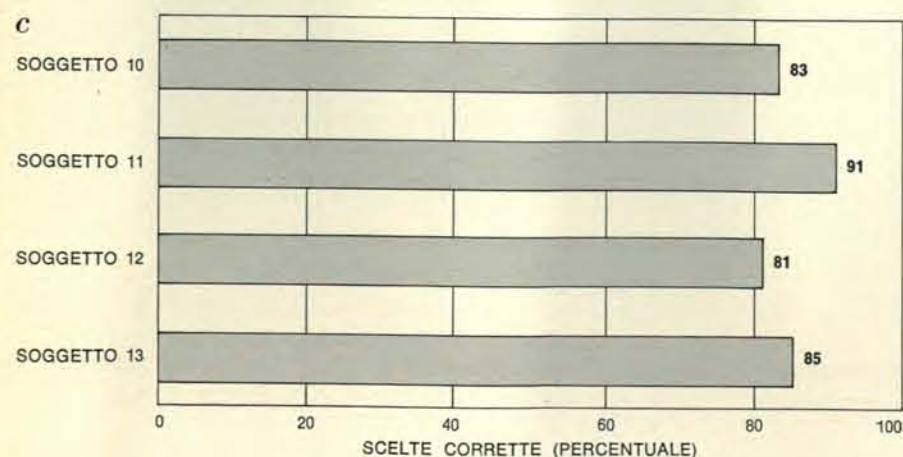
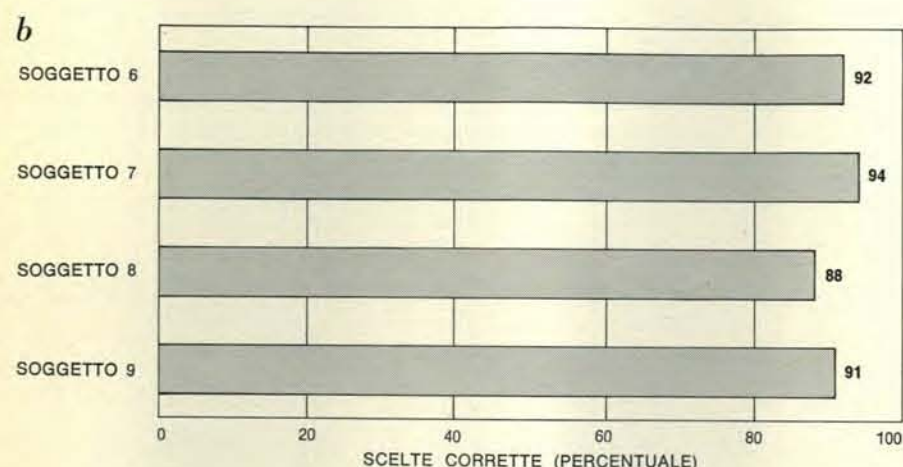
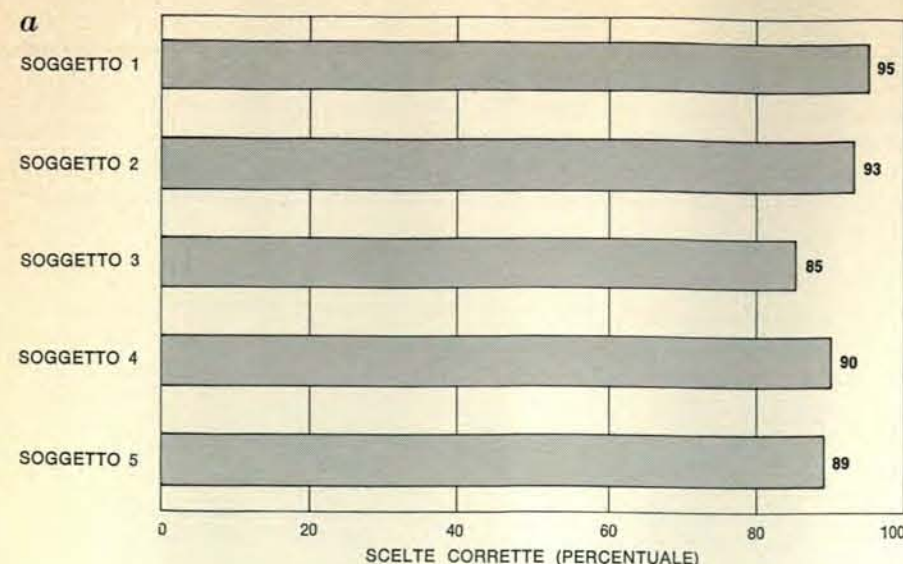
Dopo queste domande iniziali a metà dei soggetti si fece praticare individualmente un esercizio di tiro a segno, che venne presentato come un esperimento a sé. A tutti gli altri soggetti si chiese di starsene comodamente seduti, guardando uno schermo, rilassandosi e pronunciando qualunque parola venisse loro in mente. Le prime dieci parole pronunciate dai soggetti furono scritte ciascuna su un cartoncino; poi si porsero ai soggetti i cartoncini, uno per volta, chiedendo loro dapprima di associarvi ulteriori parole, poi di esprimere qualunque pensiero venisse loro in mente in connessione con la parola scritta. L'intero esercizio di associazione durava di solito circa 30 minuti.

Dopo il tiro a segno o il compito di associazione, si chiese di nuovo ad ogni soggetto di cercar di ricordare la scena, descrivendola a parole e disegnandola un'altra volta. Si posero loro le stesse domande stimolanti. A tutti i soggetti fu addotta una giustificazione per il compito interpolato; a quelli che si erano esercitati al tiro a segno si disse che la memoria avrebbe dovuto migliorare, perché avevano trascorso 30 minuti pensando a cose che non avevano nulla a che fare con il compito; a quelli che avevano associato si disse che la memoria avrebbe dovuto migliorare perché avevano trascorso 30 minuti esercitandola intensamente.

Risultò che in quelli che avevano praticato il tiro a segno la capacità di ricordare i particolari delle immagini non era né migliorata né peggiorata; invece tutti i soggetti che avevano trascorso l'intervallo associando recuperarono alcuni particolari che erano sfuggiti nella rievocazione precedente. Analizzammo anche il contenuto delle associazioni stesse, trovando che se nell'associazione era preminente un particolare che prima non era stato rievocato, era piú probabile che il soggetto lo recuperasse nella rievocazione successiva. Questi risultati, assieme a quelli di altre parti dell'esperimento, indicano che nella memoria si conservano alcune informazioni sui particolari secondari, anche se normalmente il soggetto non è in grado di rievocarli. Se cosí non fosse, anche con i piú intensi esercizi di memoria e di libera associazione non sarebbero affiorati piú particolari di quelli che erano stati originariamente rievocati.

Da questo esperimento si può trarre anche un'altra conclusione: sembrerebbe che le immagini originariamente non si depositino nella memoria sotto forma di parole, perché, se cosí fosse, i particolari potrebbero essere rievocati alla prima interrogazione; invece è necessario un periodo di intensa attività associativa, in cui il soggetto abbia la possibilità di apporre delle parole alle immagini visive, il che gli permette poi di rievocare i singoli particolari dell'immagine.

Il primo di questi esperimenti con stimoli visivi fa pensare che le immagini si conservino nella memoria in misura praticamente totale: i risultati probabilmente sarebbero stati uguali se avessimo usato 25 000 immagini invece che 2500. Il secondo esperimento indica che questo tipo di riconoscimento si fonda su un certo genere di rappresentazione mnemonica che si mantiene senza denominazioni o parole, e senza bisogno di essere formulato verbalmente. Se la rappresentazione avesse carattere linguistico, i soggetti richiesti di ricordare i particolari di un'immagine in parole o altri simboli ne ricorderebbero molti di piú di quanto in effetti avvenga. I risultati della prova fanno anche pensare che, dato che le immagini non si depositano in parole, non possano neppure essere ricordate in parole, o per lo meno non

Per l'esperimento di memoria visiva occorreva che un soggetto osservasse 280 coppie di diapositive; un'immagine di ogni coppia gli era già stata mostrata precedentemente in una serie di 2560 immagini al ritmo di una ogni 10 secondi. Qui il soggetto schiaccia un bottone per segnalare che l'immagine di sinistra è una già vista. I soggetti ricordano quasi tutte le immagini.

I risultati dell'esperimento di memoria visiva indicano che la capacità di ricordare immagini può essere anche illimitata. I soggetti in *a* riconobbero l'85-90 per cento delle 2560 diapositive viste in precedenza. Per determinare se la stanchezza ha un effetto riduttivo sulla capacità mnemonica, i soggetti 4 e 5 osservarono 1280 diapositive al giorno per due giorni, mentre i soggetti 1, 2 e 3 ne videro solo 640 al giorno per quattro giorni. Si constatò con sorpresa che non c'era una differenza significativa fra i punteggi dei due gruppi. I soggetti in *b* guardarono le diapositive presentate loro secondo l'immagine speculare, in modo che la parte destra veniva a trovarsi a sinistra e viceversa, e tuttavia riuscirono a identificare le immagini non meno precisamente dei soggetti in *a*. I soggetti in *c* videro anch'essi immagini speculari, e quando si chiese loro se l'orientamento delle fotografie era cambiato risposero in genere correttamente.

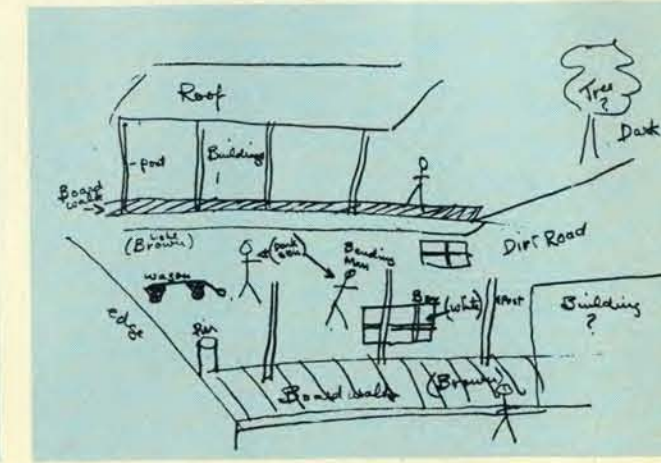
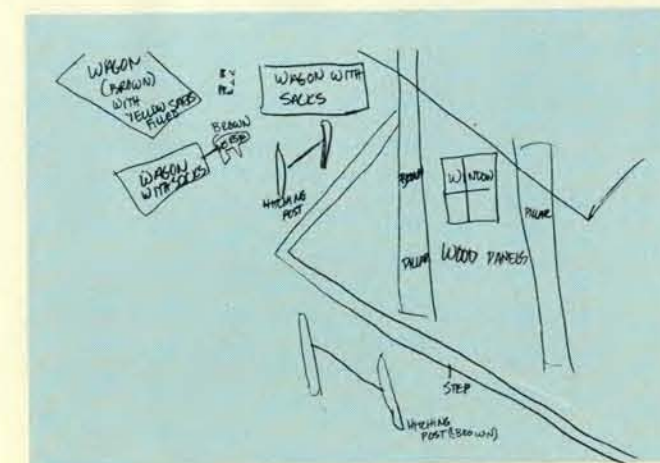
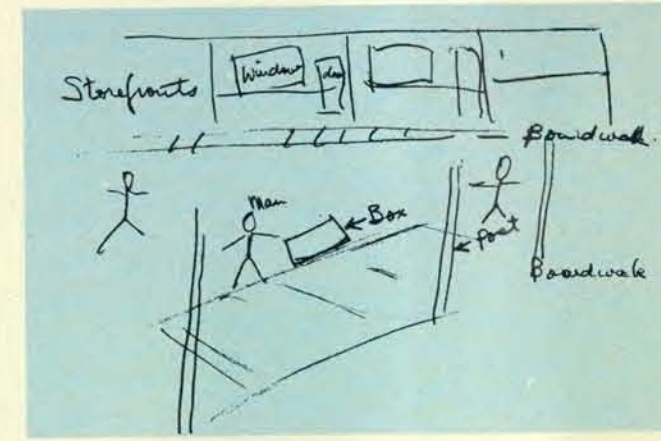
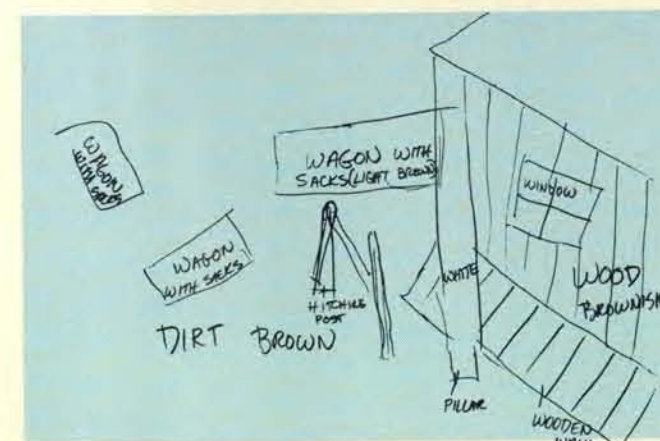
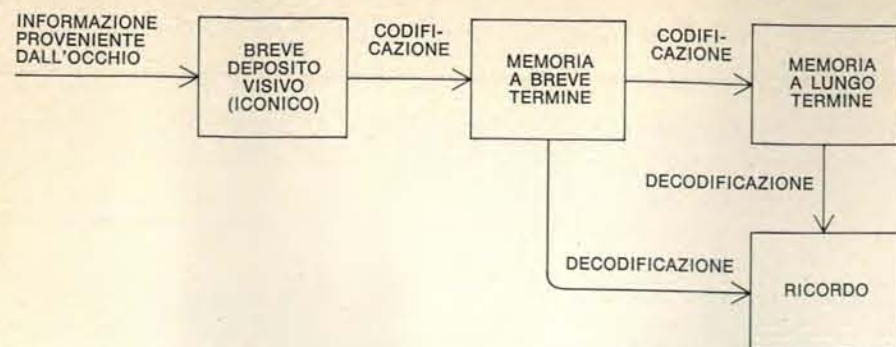
molto particolareggiatamente, a meno che la memoria non venga stimolata da un'attività quale l'esercizio di libera associazione.

Un'implicazione di tali risultati è che, se si trovasse una tecnica per facilitare l'apposizione di parole alle immagini visive, la memoria potrebbe straordinariamente migliorare. Alcune persone ritengono di possedere questa capacità, per esempio certi uomini politici, che sembrano capaci di associare un nome ad ogni viso incontrato in passato. Secondo Freud la libera associazione costituisce un modo ideale per recuperare ricordi apparentemente perduti, e benché Freud si occupasse dei ricordi rimossi piuttosto che di quelli meramente inevocabili, si può dire in generale che il ricordo di una stimolazione precedente può venir meno perché non ci serviamo di parole per ricordare immagini o sentimenti, e quindi ci è difficile usare le parole per descrivere poi il ricordo.

Se si confronta il processo della memoria figurativa con quello della memoria di parole, numeri e altri simboli percepiti visivamente, si vede che i due tipi di memoria sono probabilmente assai diversi. Entrambi trattano materiale che viene percepito quando la luce stimola la retina, generando impulsi che sono poi codificati, organizzati e inviati al cervello. Nel caso della memoria figurativa, però l'immagine viene ricevuta e si deposita permanentemente sotto forma visiva; quando invece si ha a che fare con parole o altri simboli, il primo passo della memoria consiste nello spogliare lo stimolo dalla sua forma visiva, codificandone gli elementi ed estraendone il significato. L'insieme di lettere che costituiscono una parola stampata non viene ricordato come l'immagine di diverse lettere distinte su di una pagina, ma si deposita e viene ricordato come parola; le parole vengono ricordate come idea, non come collezioni letterali di vocaboli; un segnale stradale non viene ricordato come un pannello dai colori vivaci con su una freccia o un avvertimento, ma come il messaggio di fermarsi, di rallentare o di voltare.

Questo particolare tipo di processo mnemonico spiega la facilità con cui a un lettore possono sfuggire gli errori ortografici in un testo stampato: invece di visualizzare la parola come essa appare nella realtà fisica, il lettore tende quasi immediatamente a estrarre la parola stessa dei caratteri stampati, e quindi non vede l'errore. Similmente, a un correttore di bozze inesperto può sfuggire che una singola lettera di una parola è stampata in carattere diverso dalle altre, poiché non vede la forma fisica della lettera ma l'ortografia della parola.

Sembra che il processo per estrarre il materiale linguistico dalla sua forma figurativa e depositarlo sotto forma concettuale consista di diversi passi (*si veda l'illustrazione a pag. 74 in alto*). Il primo passo è un breve momento di deposito « iconico » o « visivo »: come vedremo, in questa fase, che dura meno di un secondo dopo la percezione, l'immagine può essere vagliata e codificata. (Per esempio, una parola può essere estratta dalla forma di aggregato di lettere e tradotta in forma parlata.) Questo elemento viene poi codificato nella memoria a breve termine, e quindi passato in quella a lungo termine.

La memoria a breve termine può

Il ricordo dei particolari indica che le figure, le facce e altro materiale figurativo si depositano nella memoria sotto forma di immagini. In questo esperimento si mostrò a due soggetti la stessa fotografia di una scena campestre (*immagini in alto*), chiedendo ad ognuno di ricordare quanti più elementi poteva dell'immagine descrivendola a parole e disegnandola (*immagine centrale*). A questo punto i due soggetti ricordarono circa la stessa quantità di elementi. Poi uno dei due soggetti si esercitò al tiro a segno per mezz'ora, mentre l'altro cercò di ricordare ulteriori particolari pensando all'immagine e pronunciando qualunque parola gli veniva in mente. Dopo si chiese a entrambi i soggetti di disegnare di nuovo la scena. Il disegno in basso a sinistra, quello del soggetto che aveva praticato il tiro a segno, mostra che la sua memoria era rimasta relativamente immutata dopo l'esercizio. Il disegno dell'altro soggetto, (*in basso a destra*), contiene molti particolari che non erano presenti nel suo disegno precedente, il che indica che l'esercizio mnemonico lo aveva aiutato ad apporre le parole ai particolari visivi nella memoria. Risulta importante notare che se tali particolari si fossero depositati nella memoria sotto forma di parole anziché di immagini, entrambi i soggetti li avrebbero ricordati tutti uniti alla prima rievocazione.

Il processo mnemonico per quel che riguarda le parole, i numeri, e altro materiale linguistico passa attraverso diversi stadi. A sinistra un'informazione vista dall'occhio, come per esempio una parola stampata, depositata per breve tempo nel modo in cui appare in realtà. Mentre questa immagine visiva persiste, la parola può essere vagliata, anche se lo stimolo originario è scomparso. Al centro, durante il vaglio, la parola viene codificata nel nome corrispondente (le lettere *c, a, n, e* possono divenire la parola « cane ») e viene immessa quindi nella memoria a breve termine. A destra la parola entra nella memoria a lungo termine, forse attraverso una nuova codificazione, che questa volta estrae dal suo nome il significato. A destra in basso infine la parola risulta ricordata dal soggetto grazie alla decodificazione nel nome corrispondente.

trattenere probabilmente senza decadimento da quattro a sei elementi non collegati, ma, oltre a questo numero, occorre una forma di enunciazione per evitare che il materiale vada perduto. Un esempio comune è la necessità di ripetere un numero telefonico che abbiamo appena guardato o udito mentre corriamo al telefono per girarlo: le sette cifre non possono sopravvivere nella memoria piú di pochi secondi se non vengono ripetute.

Un altro espediente per aumentare la capacità della memoria a breve termine consiste nel ricodificare gli elementi da ricordare: una lunga serie di lettere può ad esempio essere trattenuta piú facilmente nella memoria a breve termine se viene composta in una sigla o in una parola, e in tal caso viene ricordata la parola piuttosto che le sue singole lettere.

Come avviene il passaggio del materiale dalla memoria a breve termine a quella a lungo termine? Probabilmente questo passaggio è sotteso da una ricodificazione dei nomi degli elementi nei loro significati, cioè da una codificazione semantica. A questo punto occorre un ulteriore mantenimento per trattenere l'informazione in forma rievocabile, anche se può essere necessario che l'informazione sia ricodificata sotto forma di nomi per poter essere rievocata. Non sappiamo ancora se la consapevolezza di vedere lo stimolo si identifica col deposito iconico nella memoria a breve termine, se invece si tratta di due processi simultanei, oppure se la consapevolezza è successiva alla codificazione semantica.

L'immagine visiva che persiste nella memoria di un osservatore fa sembrare che un lampo di luce persista dopo che in realtà si è estinto. In un esperimento volto a dimostrare questa persistenza, un soggetto osservò una serie di lampi, che duravano ognuno fra i 10 e i 15 millisecondi ed erano distanziati di 250 millisecondi. In *a* i lampi sono rappresentati da punte. In *b* le immagini suscitate dai lampi sono rappresentate mediante curve indicanti come i lampi apparivano al soggetto. Le curve mostrano che ogni lampo suscita un'immagine visiva (*fase ascendente rapida*). Questa immagine persiste nella memoria, svanendo gradualmente in 250 millisecondi, in modo da durare piú che lo stesso lampo. Poiché un'immagine è ancora percepibile quando inizia la immagine successiva (*intersezione delle curve discendenti e ascendenti*), al soggetto sembra che i lampi si fondano in una serie a sfarfallio (*flichering*).

Questo è un modello abbastanza generalizzato del processo della memoria linguistica; sono stati proposti anche molti modelli specifici, e i dati sperimentali non sono ancora abbastanza completi per permettere di scegliere fra essi. Tuttavia diversi recenti esperimenti hanno chiarito ciò che accade nei primi stadi; fra l'altro si è studiata la fonte degli errori in cui un individuo ricorre quando gli si chiede di ricordare diversi elementi di una vasta serie. Queste omissioni indicano una limitata capacità mnemonica, oppure dipendono dalla mancata percezione degli elementi?

È noto che, senza l'aiuto della verbalizzazione o della ricodificazione, si possono ricordare solo da quattro a sei elementi. Sembra quindi che, anche se forse gli oggetti sono stati inizialmente percepiti, alcuni vadano perduti perché nella memoria a breve termine non c'era posto per loro. Gli elementi potrebbero anche andar perduti in stadi successivi del processo mnemonico.

Come si possono confermare queste ipotesi? Come, in altri termini, si può distinguere sul piano sperimentale la mancata percezione da una limitazione della capacità mnemonica? Non si può risolvere il problema limitandosi a chiedere al soggetto se all'inizio aveva percepito ogni elemento presentatogli: se la capacità mnemonica di un soggetto è troppo limitata per contenere tutti gli elementi presentatigli, egli non sarà in grado di ricordare se inizialmente li aveva percepiti tutti.

George Sperling, dei Bell Telephone Laboratories, ha riscoperto un metodo sperimentale per affrontare questo problema, escogitato un secolo fa da N. Baxt. In uno degli esperimenti di Sperling fondati sul metodo di Baxt si presentò al soggetto una serie di lettere al tachistoscopio. La serie rimaneva visibile per circa 50 millisecondi. Finita la presentazione, si chiedeva al soggetto di ricordare tutte le lettere, finché non comparisse un segno indicante la posizione delle lettere (da uno a quattro) che egli avrebbe dovuto rievocare. Se non c'era ritardo, se cioè l'indicatore compariva immediatamente dopo la fine della presentazione, il soggetto praticamente non faceva alcun errore. Il periodo in cui egli poteva riferire le lettere senza errori durava circa 50 millisecondi (un quarto di secondo). Con l'aumentare dell'intervallo fra la fine della presentazione e la comparsa dell'indicatore aumentava invece anche la frequenza degli errori.

Da questi risultati si possono trarre alcune importanti conclusioni sul processo mnemonico e sulla fonte degli errori: se il ricordo è perfetto nell'istante immediatamente successivo all'esposizione tachistoscopica, ne consegue che probabilmente quasi tutte le informazioni presentate si sono depositate nella memoria. Infatti, se le informazioni non si fossero depositate in misura praticamente totale, il soggetto non sarebbe in grado di ricordare alcuni elementi, dato che quelli che gli si chiede di rievocare sono scelti a caso, e non sarebbe quindi in alcun modo possibile che il soggetto rievocasse solo gli elementi richiestigli.

Data la precisione del ricordo, si può anche concludere che la percezione iniziale, cioè l'immagine trasmessa dal sistema visivo al sistema mnemonico, deve anch'essa essere assai fedele: se la percezione non fosse particolareggiata ed esauriente, l'immagine iconica depositata nella memoria conterrebbe errori, e cosí pure la rievocazione. Quindi la percezione iniziale non può essere la fonte degli errori.

Poiché non si chiede mai al soggetto di rievocare piú di quattro elementi, una quantità cioè che rientra nella capacità della memoria a breve termine, gli errori e le omissioni compiuti dai soggetti quando cercano di rievocare gli elementi di una serie numerosa devono avere origine negli stadi di processuali ulteriori della memoria.

Questi esperimenti fecero pensare, ma non confermarono direttamente, che l'immagine iconica persista nella memoria per circa un quarto di secondo. Standing ed io escogitammo altri due esperimenti per averne la dimostrazione, e anche i nostri risultati confermano che l'immagine iconica ha carattere figurativo. Supponemmo che, se una rappresentazione visiva persiste per circa 250 millisecondi dopo la cessazione dello stimolo, un soggetto dovrebbe poter valutare direttamente l'entità di questa persistenza. Sapevamo però che non è possibile chiedere a un soggetto di valutare in modo assoluto tale durata: con un opportuno adattamento di intensità, i lampi che durano da un nanosecondo (10^{-9} secondi) a 10 millisecondi (10^{-2} secondi) sembrano uguali. Tentammo perciò con due altri procedimenti, entrambi notevolmente piú diretti.

Il nostro primo esperimento doveva convalidare l'assunto che una breve presentazione crea un'immagine iconica persistente nella memoria per circa

Il tachistoscopio (*scatola grande a destra*) è una camera oscura dove un soggetto osserva stimoli in rapido mutamento, come immagini o lampi di luce. L'apparato a sinistra controlla il tempo di presentazione degli stimoli. Il soggetto ha anche una cuffia acustica nella quale controlla il generatore sonoro che produce il clic in modo che un clic coincida con la comparsa di un'immagine e l'altro con la sua scomparsa. Questi esperimenti rivelano come la memoria elabora lettere, numeri e materiale simile.

250 millisecondi dopo la cessazione dello stimolo. Cosí stando le cose, si poteva prevedere che, se l'intervallo fra i lampi è uguale o poco inferiore alla durata dell'immagine iconica, il soggetto dovrebbe aver l'impressione che nessun lampo si è completamente estinto prima dell'inizio di quello nuovo: la serie delle stimolazioni dovrebbe aver carattere a sfarfallio (*flichering*), ma non ci dovrebbero essere intervalli completamente scuri fra i lampi. Sottoponemmo a verifica questa previsione presentando per breve tempo un piccolo cerchio nero su bianco in un canale di un tachistoscopio a due canali, l'altro canale era portato alla stessa illuminazione di quello che presentava il cerchio, per mantenere gli occhi del soggetto adattati alla luce. Il lampo che presentava il cerchio durava da 10 a 15 millisecondi, e si presentarono alcune serie di lampi, ognuna con un diverso intervallo fra i lampi. Dopo che aveva visto una particolare serie di lampi, si chiedeva al soggetto di indicare se il cerchio era completamente svanito ogni volta prima di ricomparire: finché l'intervallo fra i lampi non superava i 250 millisecondi circa, tutti i soggetti riferirono che la forma non era mai completamente svanita; a intervalli superiori a 250 millisecondi i soggetti riferiscono che il cerchio era svanito completamente.

La durata del lampo stesso aveva un effetto, e quale, sulla durata dell'immagine iconica? Riscontrammo che per tempi relativamente brevi di presentazione, situati fra i 4 e 200 millisecondi (molto al di sopra della soglia), l'immagine iconica persisteva per circa 250 millisecondi. In altri termini, la durata dell'immagine iconica sembrava indipendente dalla durata dello stimolo originario, per lo meno per stimoli molto brevi.

La conferma dell'esistenza di un'immagine iconica sembrava dunque rag-

giunta: tuttavia nessuna delle prove eseguite finora escludeva la possibilità che si fossero misurate le proprietà di un'immagine retinica piuttosto che di un'immagine formatasi durante i primi stadi del processo mnemonico. Per eliminare tale possibilità (o per confermarla) escogitammo un altro procedimento. In questa variante dell'esperimento il primo lampo di una serie veniva presentato all'occhio destro, mentre si bloccava la visione dell'occhio sinistro; poi si bloccava la visione dell'occhio sinistro, presentando il lampo successivo all'occhio destro, e via dicendo. In tali condizioni ai soggetti occorrevano ancora 250 millisecondi fra un lampo e l'altro. Questo risultato indica chiaramente che, anche se ci può essere una certa persistenza a livello retinico, l'immagine iconica esiste a livello centrale, dopo che si sono combinate le informazioni provenienti dai due occhi. Se cosí non fosse, sarebbe bastato che i lampi si distanziassero di soli 125 millisecondi (250 millisecondi per ogni occhio separatamente).

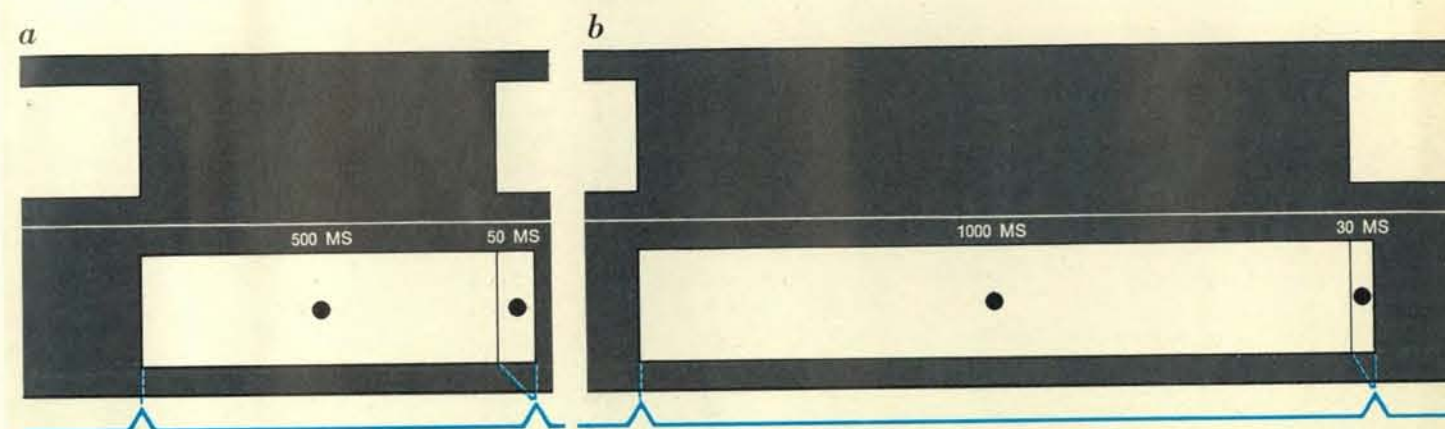
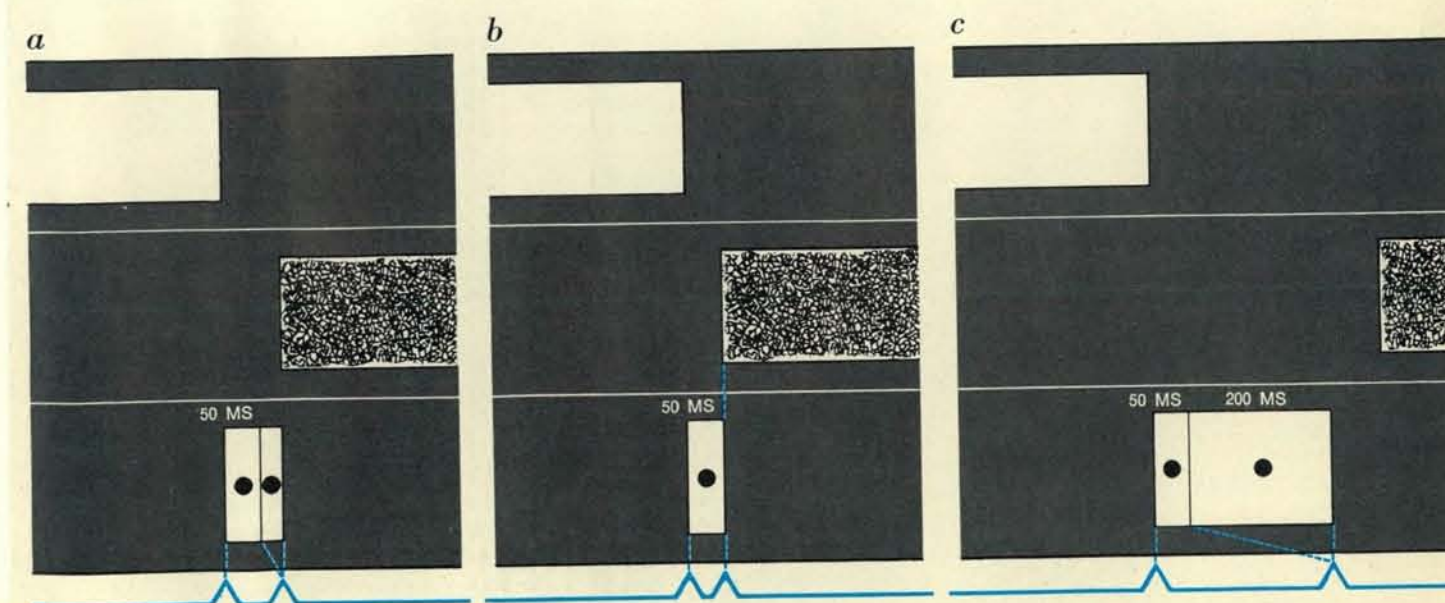
Nell'esperimento successivo cercammo di perfezionare ed estendere i risultati raggiunti. Seguendo un piano proposto da Sperling, ci servimmo di un tachistoscopio a tre canali. Un canale era costituito da un campo vuoto illuminato, il secondo presentava l'immagine di cui il soggetto doveva stimare la durata, il terzo presentava disegni casuali (l'equivalente visivo del « rumore »). Il soggetto aveva una cuffia acustica attraverso la quale udiva un breve clic circa al momento in cui gli veniva presentata l'immagine al tachistoscopio. Egli aveva il compito di girare una manopola, regolandola in modo che il clic e l'inizio del lampo coincidesse esattamente. Si presentavano il clic e il lampo ogni pochi secondi. Quando un soggetto era soddisfatto della corrispondenza fra clic e inizio della presentazione, si ripeteva il procedimento, e questa volta egli doveva far coincidere un altro clic con la fine della presentazione. L'intervallo di tempo fra il primo e il secondo clic rappresentava dunque la sua valutazione della durata della presentazione.

Come avevano previsto, il secondo clic seguiva di circa 200 millisecondi la reale cessazione della brevissima esposizione, il che indica che l'immagine iconica ha questa durata. Si ricorderà che nell'esperimento precedente la durata dell'immagine iconica era indipendente dalla durata di un lampo inferiore ai 200 millisecondi. Invece in questo esperimento la persistenza dell'immagine diminuí con l'aumentare della durata dello stimolo: quando l'esposizione durava piú di 500 millisecondi, l'immagine persisteva meno di 50 millisecondi; l'immagine iconica associata a un lampo di un secondo è inferiore ai 30 millisecondi; l'immagine iconica dura piú di circa 250 millisecondi (400 millisecondi) solo quando il soggetto è adattato all'oscurità.

Entrambe queste ricerche concordano strettamente con le valutazioni piú complicate che si trovano nella letteratura, e confermano che nel processo mnemonico sembra aumentare la durata di tutti i lampi brevi, fino ad almeno 200 millisecondi; se i lampi hanno già questa durata, gli aumenti ulteriori si riducono al minimo. Queste ricerche inoltre dimostrano chiaramente che la memoria a breve termine postulata da Sperling ha in effetti carattere visivo, poiché i soggetti descrivono l'immagine depositatasi come se parlassero del lampo stesso.

Quali fattori determinano la durata dell'immagine visiva? Essa può essere allungata o abbreviata in circostanze normali o sperimentali? Che funzione svolge nell'estrazione delle informazioni dagli stimoli visivi? Dovrebbe essere evidente che una breve persistenza dell'immagine visiva, che prolunga lo stimolo per meno di un secondo, risulterà utile solo se tale prolungamento fornisce un certo vantaggio critico. Forse gli individui sono chiamati a descrivere gli effetti di brevi lampi solo in situazioni sperimentali? Forse abbiamo inventato un concetto di persistenza visiva che non ha alcuna funzione in natura (eccetto forse di permetterci di leggere al buio durante i lampi di un temporale)? Io sono fermamente convinto del contrario. Nell'ambiente si incontrano molti stimoli visivi che compaiono solo per un tempo brevissimo. Per gli adulti colti l'esperienza visiva piú importante e continuativa è la lettura, e la lettura non è altro che il compito di passare in rassegna una rapida successione di brevi esposizioni visive che possiedono un altro contenuto di informazione.

Molti studiosi hanno messo in luce che la persistenza dell'immagine iconica è di aiuto nella lettura, e quindi anche in altri simili compiti visivi. Tali studiosi hanno riscontrato che una persona che non sappia leggere velocemente, o che voglia leggere lentamente, fissa gli occhi su ogni parola per circa un quarto di secondo. Anche un lettore veloce fissa gli occhi per circa lo stesso lasso di tempo. La maggior velocità di lettura si ottiene riducendo il numero delle fissazioni, cioè

Misurazione della durata dell'immagine visiva. In un esperimento qui schematicamente rappresentato si chiese a un soggetto di far corrispondere il clic (*in basso*) alla comparsa di uno stimolo ripetitivo consistente in un cerchio nero in campo bianco mostratogli al tachistoscopio. Il campo di adattamento nel tachistoscopio (*in alto*) tiene adattati gli occhi del soggetto a un determinato livello di luce. In *a* il soggetto ode un clic dopo aver visto lo stimolo; in *b* ha cominciato a far coincidere il clic con la comparsa dell'immagine successiva. In *c* è riuscito a compiere la regolazione e ha anche fatto corrispondere un secondo clic con quella che egli percepisce come la cessazione dello stimolo. In effetti il clic segna la fine dell'immagine visiva nella memoria del soggetto, piuttosto che la fine dello stimolo stesso, che è svanito 200 millisecondi prima. L'intervallo fra la comparsa dello stimolo e il clic rappresenta dunque la durata dell'immagine visiva.

Uno stimolo di lunga durata al tachistoscopio produce una immagine visiva breve nella memoria. In *a* il soggetto ha già adattato un clic in modo da farlo coincidere con l'inizio dello stimolo, e un altro in modo da farlo coincidere con la cessazione percepita. Lo stimolo in realtà dura 500 millisecondi. L'intervallo fra i clic è però di circa 550 millisecondi, il che indica che l'immagine visiva dura 50 millisecondi. In *b*, l'immagine visiva è piú breve, di circa 30 millisecondi.

L'immagine visiva è cancellata dalla comparsa di un nuovo stimolo, In *a* il cerchio nero è presentato per 50 millisecondi, un intervallo normalmente abbastanza lungo per produrre un'immagine visiva di 200 millisecondi. In questo caso però compare il rumore visivo (una configurazione casuale di lettere). Questo nuovo stimolo fa cessare l'immagine visiva del cerchio nero dopo soltanto pochi millisecondi. Quindi l'intervallo fra i clic è solo di poco piú lungo della durata dello stimolo. In *b* il rumore incomincia quando il cerchio scompare, in modo che l'intervallo fra i clic è uguale alla durata dello stimolo, e ciò indica amaramente che non c'è alcuna immagine visiva. In *c* il rumore visivo non ha effetto sull'immagine perché esso inizia in un periodo successivo, dopo che l'immagine è svanita.

guardando ed elaborando diverse parole invece che una sola durante ogni fissazione. A entrambi i tipi di lettore occorrono da 30 a 50 millisecondi per spostare gli occhi da un punto di fissazione al prossimo.

Sembrerebbe dunque che sia i lettori veloci sia quelli lenti abbiamo bisogno di circa un quarto di secondo per percepire ed estrarre le informazioni contenute in ogni singola parola. Poiché la lunghezza dell'intervallo è controllata dal lettore, si può trarne la conclusione che tale intervallo costituisca il tempo minimo per codificare il materiale linguistico nella memoria. Se ciò è vero, ne consegue che la persistenza dell'immagine visiva potrebbe servire a prolungare uno stimolo che non dura abbastanza a lungo per essere riconosciuto altrimenti.

Non si deve considerare una mera coincidenza il fatto che la natura della persistenza iconica e il tempo minimo di fissazione nella lettura sono entrambi di circa un quarto di secondo.

I dati in sostegno di questa linea esplicativa cominciano ad accumularsi, ma sono tuttora inadeguati. Le dimostrazioni riguardano in parte il modo in cui un'immagine iconica viene cancellata e gli effetti della cancellatura sui processi concernenti l'informazione. Se la persistenza della immagine iconica aiuta il soggetto a elaborare l'informazione, vi dovrebbe anche essere un modo per cancellare l'immagine iconica quando ha servito al suo scopo. Se non ci fosse un processo di cancellatura, un lettore o un individuo che rapidamente vaglia i visi di una folla potrebbe essere ostacolato dalla persistenza dell'immagine iconica di uno stimolo precedente mentre tenta di assimilarne uno nuovo.

Molti esperimenti mostrano che la sovrapposizione di una nuova forma visiva o di un campo di disturbo visivo su quello vecchio fa veramente scomparire l'immagine iconica, ostacolando o facendo cessare l'elaborazione dell'informazione contenuta nella forma. Ciò accade normalmente nella lettura, quando gli occhi si concentrano su un nuovo pezzo di stampa: lo stimolo nuovo spazza via l'immagine iconica di quello precedente, in modo da spianare la via per l'elaborazione di nuovo materiale.

L'esperimento di accoppiamento fra clic e lampo è attinente a questo problema. In un altro esperimento di quella ricerca, il terzo canale del tachistoscopio presentava un rumore visivo consistente in configurazioni casuali. Questo canale fu inserito in diversi momenti dopo il lampo: se il nuovo stimolo (rumore visivo) avesse cancellato l'immagine di quello precedente, come ci aspettavamo, allora l'intervallo fra i due clic avrebbe dovuto ridursi se il rumore visivo fosse giunto prima che l'immagine iconica avesse avuto il tempo di svanire. In particolare, se il rumore visivo coincideva con la fine del lampo, l'intervallo fra i clic non avrebbe dovuto superare quello del lampo stesso. Viceversa, se il rumore visivo fosse stato differito oltre alla normale persistenza dell'immagine iconica, non avrebbe dovuto influire sulla prestazione del soggetto. E così infatti avvenne.

In questo compito non occorreva che il soggetto estraesse informazioni: gli si chiedeva solo di valutare quanto tempo gli occorreva per estrarre la informazione. I risultati parlano in favore dell'effetto di cancellatura, cioè dell'effetto di cessazione dell'elaborazione a opera del rumore visivo.

È però ancor più convincente il fatto che, quando si toglie un'immagine e se ne presenta al soggetto un'altra prima che abbia avuto il tempo di elaborarla, egli riferisce che sa di aver visto il primo stimolo ma non ha avuto abbastanza tempo per riconoscerlo. Questo risultato si può ottenere facendo seguire alla presentazione un rumore visivo oppure un'ulteriore informazione. Se ne può concludere che il rumore visivo riduce il tempo di cui il soggetto dispone per estrarre l'informazione, non il tempo di cui dispone per percepire lo stimolo; il quarto di secondo occupato dall'immagine iconica non è necessario per vedere l'immagine ma per elaborarne il contenuto.

Questo effetto si evidenziò molto chiaramente in un esperimento di riconoscimento di una sequenza di parole. Di fronte al soggetto che guardava sullo schermo di un pannello elettroluminescente comparvero tutte le lettere di una parola in successione, ognuna nella stessa collocazione, in modo che la seconda lettera distruggeva la persistenza della prima, la terza quella della seconda, e via dicendo. Così il ritmo della presentazione regolava effettivamente il tempo di cui il soggetto disponeva per elaborare ogni lettera. Il ritmo fu fatto variare da 20 millisecondi a 300 millisecondi per lettera. Non è sorprendente che la probabilità di riconoscere ogni lettera era più alta a ritmi più lenti di presentazione. Più importante (e più rilevante ai nostri fini) era il reperto che, per ogni ritmo di presentazione, non importava che il tempo fra la comparsa di una lettera e la comparsa della successiva fosse usato interamente per presentare la lettera o che invece la lettera venisse presentata soltanto per pochi millisecondi; quando una lettera è stata vista, il tempo ulteriore di presentazione è irrilevante, purché non se ne disturbi la persistenza. Il tempo rimanente viene usato per elaborare l'informazione già ottenuta.

Infine, diversi esperimenti che risalgono a una precedente esperienza di Sperling hanno dimostrato che, quando l'intervallo fra la comparsa dello stimolo e il sopraggiungere del rumore visivo varia in modo da far variare il tempo disponibile per elaborare il contenuto dello stimolo, il tempo di cui il soggetto dispone per elaborare lo stimolo e il numero degli elementi che riesce a riconoscere sono legati da una funzione quasi lineare. In un esperimento condotto nel mio laboratorio abbiamo dimostrato che questa relazione è molto più evidente in soggetti che non si sono precedentemente esercitati in questo compito particolare, e si attenua man mano che i soggetti acquistano ulteriore esperienza. Nei primi giorni dell'esperimento sembra che occorrano al soggetto circa 10 millisecondi di tempo di elaborazione per riconoscere ogni lettera dopo che ha disposto di un certo tempo per percepire tutte le lettere e costituire l'immagine iconica. Per le immagini particolari da noi usate, questo tempo di percezione e di formazione dell'immagine era di circa 50 millisecondi. Quindi per una parola di quattro lettere occorre che intercorrano circa 90 millisecondi prima dell'arrivo del rumore visivo perché ognuna delle sue lettere possa essere correttamente riconosciuta. Però, dopo diversi giorni, per riconoscere quattro lettere occorre poco più tempo che per riconoscerne una.

È evidente che molte cose accadono nel primissimi millisecondi dopo la comparsa di uno stimolo visivo che debba essere codificato nella memoria verbale. È anche evidente che il processo mnemonico può essere considerato come un sistema riguardante l'elaborazione dell'informazione, consistente in diversi stadi e con proprie costanti temporali per l'estrazione, il decadimento, le modalità di persistenza, la suscettibilità all'interferenza o alla cancellatura, e simili. Un simile modo di considerare il processo mnemonico indurrà probabilmente gli sperimentatori a escogitare anche altri esperimenti che forniranno una comprensione generale della percezione visiva. Tale conoscenza è importante in se stessa, ma spero sia anche chiaro in quale grande misura la conoscenza dei processi implicati nella lettura e in altre attività visive sia in ultima analisi connessa alla comprensione della percezione visiva e dell'elaborazione delle informazioni.

Luce e strati monomolecolari

Il comportamento della luce viene studiato impiegando film di piú strati monomolecolari di acidi grassi e coloranti fluorescenti: questi irraggiano luce come fa un'antenna con le onde radio

di Karl H. Drexhage

La luce è la forma piú familiare di radiazione elettromagnetica. Tuttavia, alcuni aspetti dell'emissione e propagazione della luce non sono stati approfonditi quanto i problemi corrispondenti nei campi, per esempio, delle microonde o delle onde radio, a causa della cortissima lunghezza d'onda della radiazione visibile, che è molto piccola in confronto alle dimensioni degli usuali apparecchi di detezione. Le onde visibili hanno lunghezza compresa tra circa 4000 e 7000 angstrom (decimillionesimi di millimetro), mentre le lunghezze delle microonde e delle onde radio variano tra qualche millimetro e qualche centinaio di metri.

Recentemente si è resa disponibile una nuova tecnica sperimentale che consente di superare le difficoltà connesse con la piccola lunghezza d'onda della radiazione visibile. Questa tecnica, che è stata messa a punto in gran parte dai miei colleghi e da me presso l'Università di Marburg, in Germania, si basa sull'impiego di sottilissimi film di materia, costituiti da foglietti monomolecolari di acidi grassi a lunga catena lineare e di coloranti fluorescenti, sovrapposti in piú strati. Tra le altre cose, questa ricerca ha consentito di scoprire che i coloranti fluorescenti, quando emettono radiazione visibile in seguito a eccitazione, si comportano in modo molto simile a un'antenna che emette onde radio.

Quando cominciammo a lavorare intorno a questo programma di ricerca, per prima cosa adattammo ai nostri fini una tecnica descritta piú di 30 anni fa da Katharine B. Blodgett e Irving Langmuir dei laboratori di ricerca della General Electric. Da allora abbiamo continuamente migliorato il loro metodo di preparazione di strutture molecolari costituite da piú strati monomolecolari sovrapposti di acidi grassi e abbiamo reso possibile l'applicazione di questa tecnica alla preparazione di multistrati costituiti da diverse altre sostanze. Prima di spiegare come questi multistrati possono essere impiegati per lo studio delle onde visibili, descriverò brevemente il metodo usato per fabbricarli.

Gli acidi grassi hanno una molecola costituita da una catena idrocarburica idrofoba (priva cioè di affinità per l'acqua) recante a una estremità un gruppo funzionale acido (carbossile) idrofilo (dotato di notevole affinità per l'acqua). Deponendo una goccia di soluzione di un acido grasso in benzolo sulla superficie dell'acqua contenuta in una vaschetta, il solvente evapora rapidamente, lasciando le molecole dell'acido allineate sulla superficie con il gruppo idrofilo immerso nell'acqua (*si vedano le figure alle pagine 84-85*). Con una quantità sufficiente di soluzione, è possibile occupare l'intera superficie libera dell'acqua con uno strato monomolecolare di acido grasso. A questo punto le gocce di soluzione non si spandono piú, ma galleggiano inalterate. Se ora si aumenta la superficie libera, per esempio muovendo una barriera galleggiante, le gocce riprendono a spandersi fino a sparire completamente o fino a che tutta la superficie è stata occupata dal film monomolecolare.

Successivamente si applica una piccola forza alla barriera galleggiante, per esercitare una pressione laterale sul film monomolecolare, cosí da spingere le molecole l'una contro l'altra in un insieme ordinato, con le catene idrofobe allineate verticalmente in una struttura compatta. Si ottiene cosí un film non piú comprimibile e sufficientemente viscoso da non poter entrare negli interstizi tra il galleggiante e le pareti della vaschetta. Lo spessore del film è pari alla lunghezza delle molecole dell'acido. Di solito impiegavamo acido arachico, ($C_{20}H_{40}O_2$) una sostanza che si estrae dall'olio di arachidi e ha una catena lineare costituita da 20 atomi di carbonio. Normalmente l'acido viene parzialmente convertito nel sale di cadmio (CdC_{20}) aggiungendo una piccola quantità di cloruro di cadmio ($CdCl_2$) all'acqua della vaschetta, al fine di migliorare la stabilità degli strati.

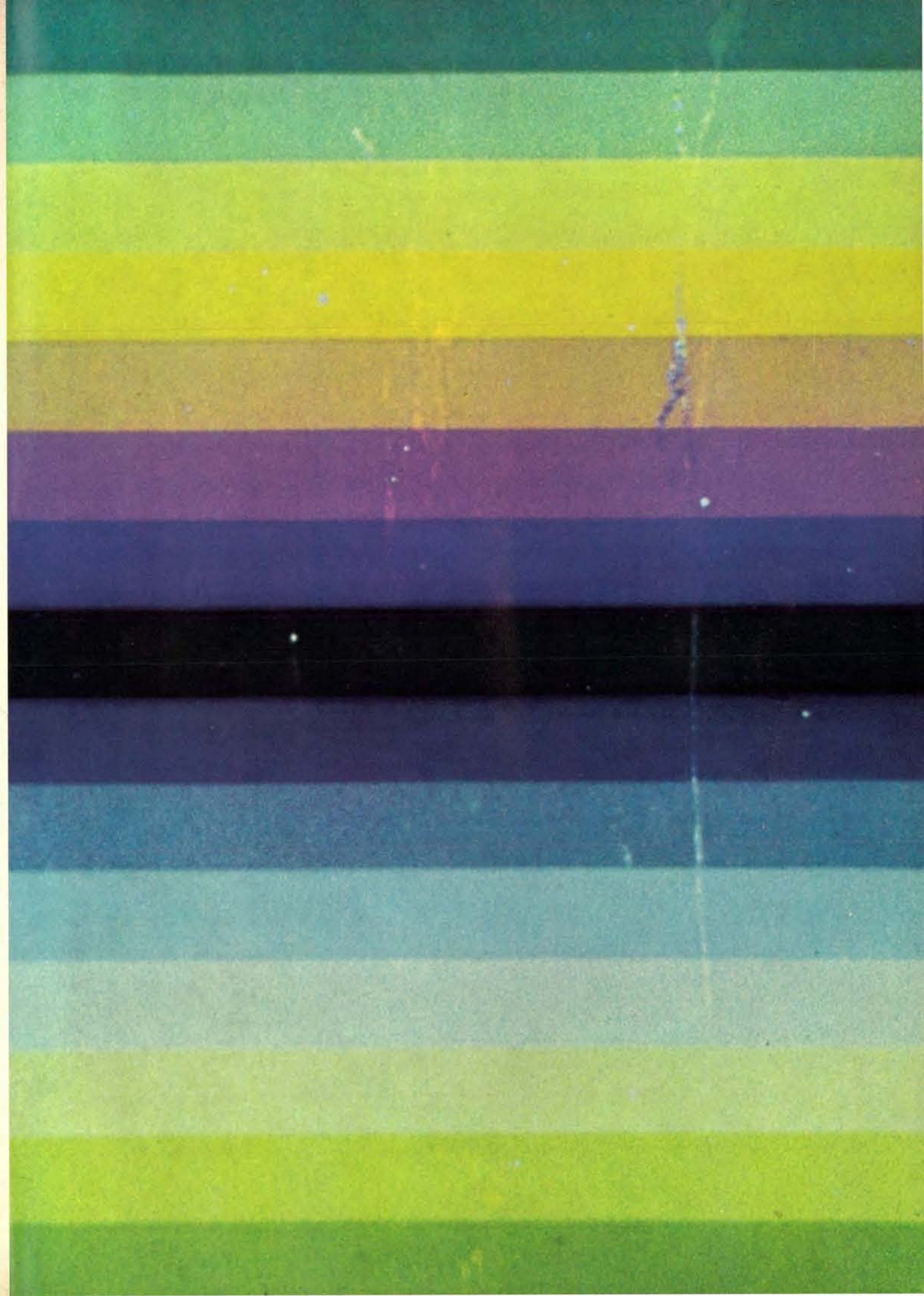
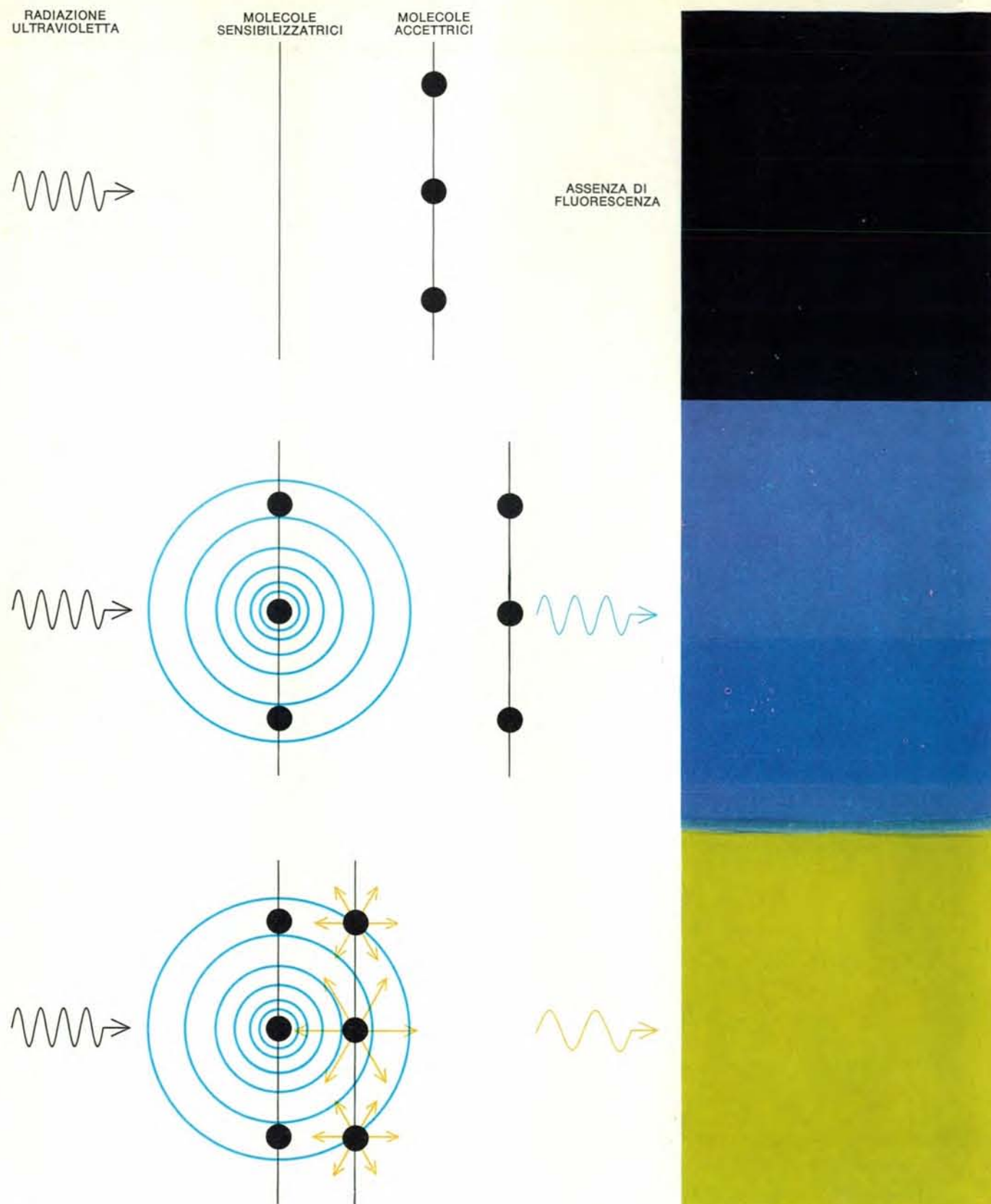
Immergendo una lastra di vetro attraverso la superficie d'acqua ricoperta dello strato monomolecolare, quest'ultimo non aderisce al vetro con le catene idrofobe. Invece, estraendo la lastra, i gruppi carbossilici idrofili aderiscono alla superficie bagnata del vetro ed entrambi i lati della lastra vengono ricoperti da uno strato monomo-

La fotografia dimostra come si può studiare l'onda stazionaria che si forma di fronte a uno specchio mediante uno strato monomolecolare di una sostanza fluorescente, deposto su una successione a gradini di strati monomolecolari di acido grasso. La distanza tra lo strato fluorescente e lo specchio aumenta dall'alto verso il basso. L'altezza di ogni gradino è di 158 angstrom, pari allo spessore di sei strati monomolecolari di acido grasso. La lastra è stata irraggiata con radiazione ultravioletta di 3660 angstrom di lunghezza d'onda, assorbita dalle molecole di cianina in proporzione al quadrato dell'ampiezza del campo elettrico nella loro posizione (*si veda la figura a pagina 88*). La netta differenza di fluorescenza sui gradini adiacenti prova che non v'è stata diffusione del colorante. Le differenze nel colore della fluorescenza sono dovute all'interferenza con le onde riflesse dallo specchio (*si veda la figura a pagina 89 in alto*).

RADIAZIONE ULTRAVIOLETTA

MOLECOLE SENSIBILIZZATRICI

MOLECOLE ACCETTRICI

ASSENZA DI FLUORESCENZA

Un campo a corto raggio di una molecola fluorescente ha intensità decrescente con l'inverso del cubo della distanza. Le proprietà di questo campo si studiano depositando su una lastra di vetro uno strato monomolecolare di una cianina con fluorescenza blu sopra il quale, distanziato con un numero variabile di strati monomolecolari di acido grasso, si deposita uno strato di una altra cianina la cui fluorescenza gialla viene eccitata proprio dalla radiazione emessa dal primo strato di cianina. Le molecole del secondo strato fluorescente, dette accettrici, fungono quindi da detettori del campo generato dalle molecole eccitate del primo strato, dette sensibilizzatrici. Sulla lastra di cui si riporta a destra la fotografia, la distanza tra i due strati fluorescenti è zero in basso: si ha quindi un trasporto completo di energia tra i due strati e si osserva solo la fluorescenza gialla. Quando i due strati distano più di 100 angstrom, come nella parte in mezzo, le molecole accettrici non sentono più l'influenza del campo delle molecole sensibilizzatrici e, non essendovi trasporto di energia, è visibile solo la fluorescenza blu. La porzione in alto è stata ricoperta solo con lo strato accettore, che non assorbe la radiazione ultravioletta di eccitazione.

lecolare di acido. La lastra emerge dall'acqua perfettamente asciutta e la sua superficie è divenuta idrofoba, in quanto le molecole di acido sono disposte con le catene idrofobe verso l'esterno. Contemporaneamente il galleggiante si sposta, mantenendo in pressione il film galleggiante. Immergendo di nuovo la lastra, si ha il deposito di un secondo strato monomolecolare sopra il primo e riestraendola se ne deposita un terzo. Questa procedura può essere ripetuta quante volte si voglia, per ottenere un multistrato il cui spessore è determinato dallo spessore dei singoli film monomolecolari (monostrati) e dal numero di questi.

Il procedimento di deposito dei multistrati è molto semplice. Occorrono solo dei prodotti chimici puri, una bacinella con l'apposito galleggiante e un sistema di immersione delle lastre di vetro (*si veda la figura a destra*).

Nella maggior parte dei nostri esperimenti non abbiamo neppure preso precauzioni particolari contro la polvere, riuscendo ugualmente a depositare multistrati dalle caratteristiche ottiche eccellenti costituiti da più di 500 strati monomolecolari.

Apparecchio per depositare strati monomolecolari su una lastra di vetro. Consiste essenzialmente in una bacinella di Lucite larga circa 30 cm, piena d'acqua e munita di una barriera galleggiante, trascinata da un peso per mezzo della puleggia in primo piano. La sostanza da depositare viene distribuita sulla superficie dell'acqua e la piastra da ricoprire viene alzata o abbassata con il sistema a manovella che si vede a destra. Nelle prossime due pagine si spiega come avviene la deposizione dello strato.

Per verificare se i multistrati venivano effettivamente prodotti con la precisione e la riproducibilità che la descrizione precedente lascerebbe sperare, li abbiamo controllati attentamente con diversi metodi indipendenti. Ne descriverò solamente due, uno elettrico e l'altro ottico.

Se si deposita un multistrato su una lastra di vetro ricoperta da film di alluminio, ottenuto per evaporazione del metallo, e poi si deposita un secondo film di alluminio sopra il multistrato, si ottiene un condensatore, la cui capacità è inversamente proporzionale allo spessore del multistrato, che funge da dielettrico. In effetti, ponendo in grafico l'inverso della capacità del condensatore in funzione del numero degli strati ottenevamo dei tracciati rettilinei, tanto più inclinati quanto maggiore era la lunghezza della catena dell'acido grasso e quindi quanto maggiore era lo spessore del film monomolecolare. I valori di capacità erano riproducibili entro l'uno per cento.

Purtroppo questo metodo ha vari inconvenienti. Innanzitutto il calore liberato nel depositare il secondo film metallico può danneggiare il multistrato. È possibile diminuire questo rischio convertendo interamente l'acido nel sale di cadmio, che fonde a temperatura elevata e non è volatile. Ma, all'aumentare del contenuto di cadmio, il film monomolecolare diviene molto viscoso e non aderisce bene al vetro. Inoltre la capacità del condensatore dipende dalla costante dielettrica del multistrato, che è difficile misurare indipendentemente. Quindi riesce impossibile determinare accuratamente lo spessore del multistrato.

Questa misura si può fare invece molto elegantemente facendo incidere una radiazione monocromatica su un multistrato depositato su uno specchio semitrasparente. L'onda piana incidente viene riflessa più volte tra lo specchio e l'interfaccia acido-aria e ogni volta perde parte della sua energia, in quanto la riflessione non è totale e parte della radiazione si perde per trasmissione oltre lo specchio (*si veda la figura a pagina 86 in alto*). Le onde trasmesse si sommano in una risultante, la cui ampiezza dipende dallo sfasamento tra le singole onde. Questo a sua volta dipende dalla diversa lunghezza del percorso effettuato dalle onde riflesse entro il multistrato e quindi l'intensità della luce trasmessa dipende dallo spessore del multistrato.

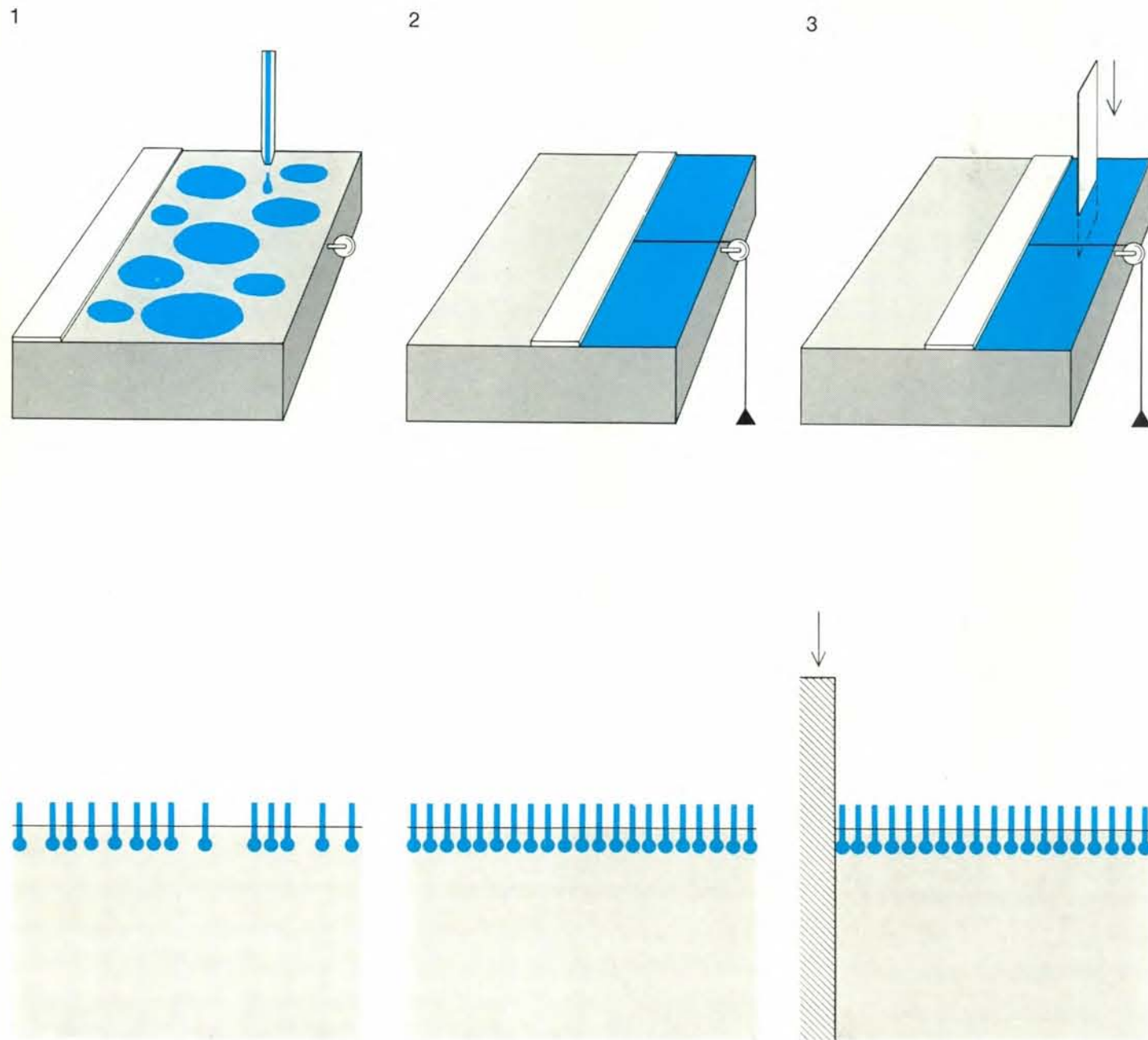
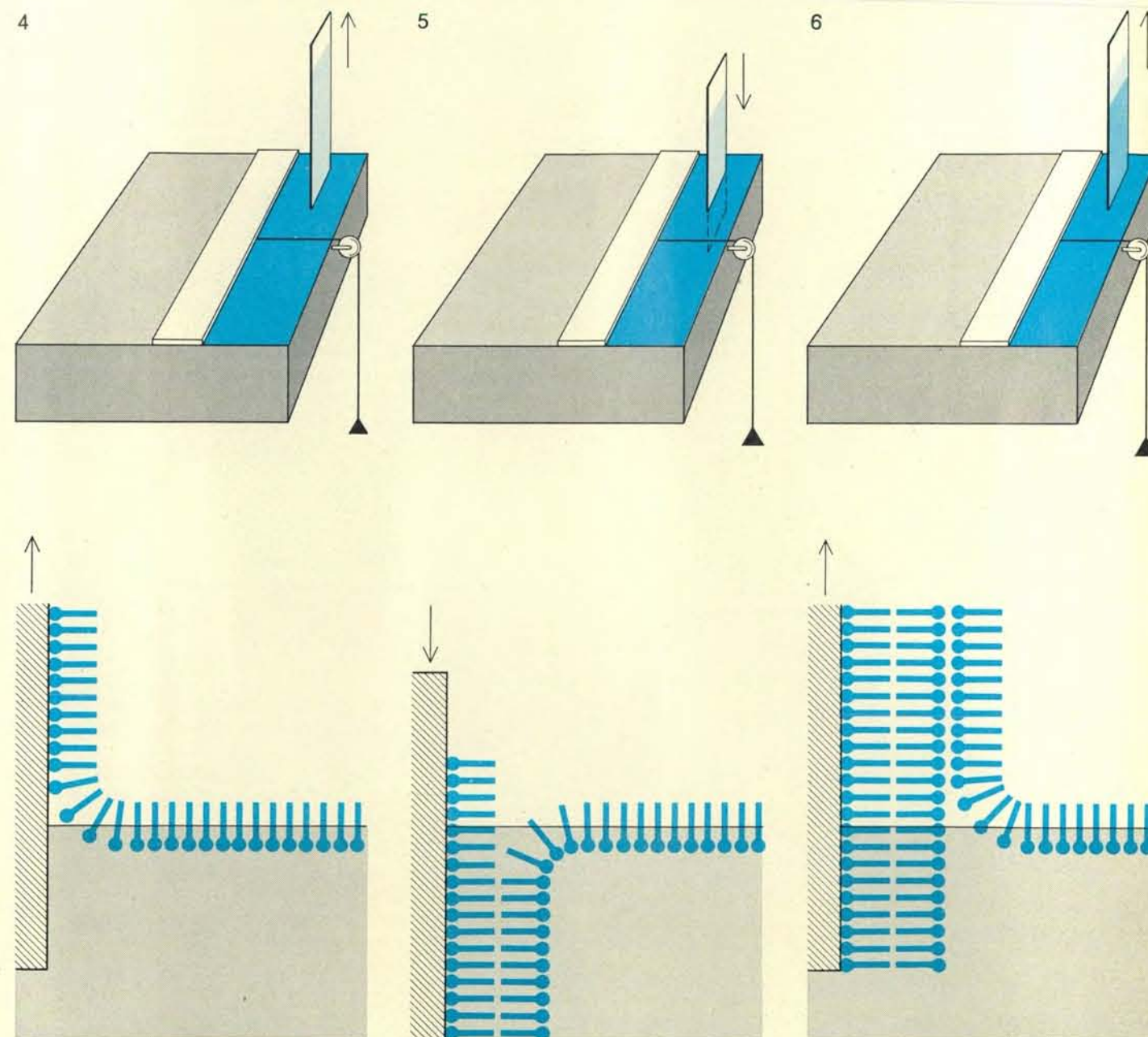
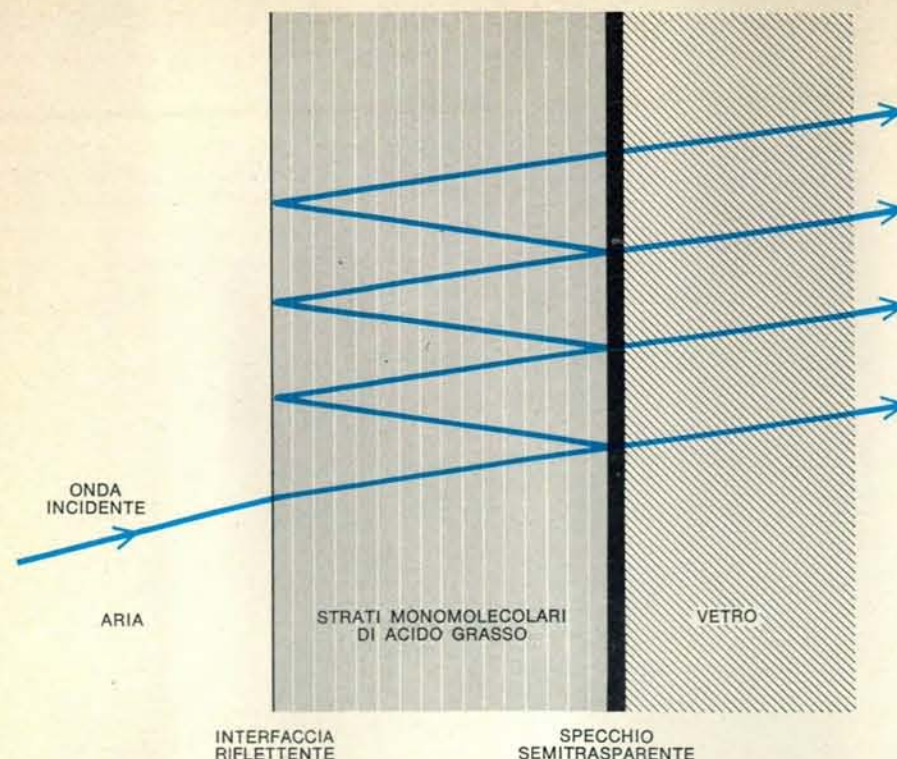
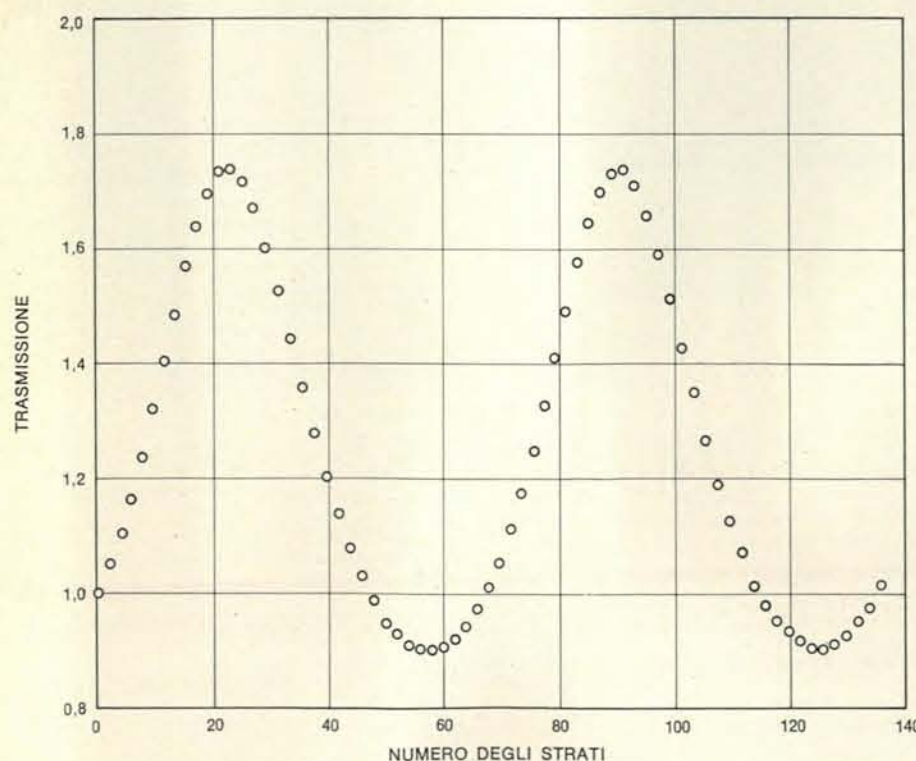
Partendo da una lastra semiargentata, vi abbiamo deposto, uno dopo l'altro, più di 100 strati di CdC_{20}, misurando ogni volta l'intensità della luce trasmessa, in confronto a una lastra identica non ricoperta (*si veda la figura a pagina 86 in basso*). I risultati mostrano, come ci attendevamo, una variazione dell'intensità della radiazione trasmessa in funzione del numero degli strati e si adattano a una curva teorica entro i limiti di errore delle nostre misure. Questo rende l'idea della eccellente riproducibilità del sistema di deposizione dei multistrati.

Con questo metodo è anche possibile misurare lo spessore degli strati monomolecolari. Il numero di strati che si debbono depositare per passare da un massimo di trasmissione all'altro corrisponde a metà lunghezza d'onda della radiazione impiegata. Questa si accorcia all'interno del multistrato in proporzione all'indice di rifrazione dell'acido, che però si può determinare indipendentemente. Da questi dati abbiamo calcolato che lo spessore di uno strato monomolecolare di CdC_{20} è di 26,4 angstrom. Con il sistema dei multistrati è quindi possibile costruire degli spessori compresi tra circa 10 e 10 000 angstrom con una precisione di poche parti per mille.

Molte altre sostanze, oltre agli acidi grassi, formano strati monomolecolari sulla superficie dell'acqua e con alcune di esse è stato possibile realizzare dei multistrati. Per i nostri esperimenti erano necessari degli strati monomolecolari di coloranti, fluorescenti, capaci cioè di assorbire e di emettere radiazione visibile, cosa dimostratasi molto difficile. I coloranti reperibili in commercio non diedero risultati soddisfacenti, in quanto non riuscivamo a impedirne la diffusione all'interno del multistrato, cosa che non permetteva l'esecuzione di diverse importanti esperienze.

Alla fine riuscimmo a sintetizzare dei coloranti fluorescenti adatti a formare monostrati indiffusibili, costituiti da un gruppo idrofilo (in questo caso chiamato cromoforo) e una lunga catena idrofobica. In effetti questi coloranti formavano strati monomolecolari sulla superficie dell'acqua, ma erano strati con scarsa resistenza meccanica, la cui superficie dipendeva dalla pressione laterale esercitata dal galleggiante e con cui non si ottenevano risultati riproducibili.

La ragione di ciò si spiega osservando la struttura molecolare di questi coloranti (*si veda la figura a pagina 87*). La sezione trasversale delle catene idrocarburiche è sensibilmente minore di quelle del cromoforo e pertanto le catene idrofobiche non possono allinearsi in una struttura compatta, come avviene invece con gli acidi grassi. Questo inconveniente venne superato aggiungendo una certa quantità di acido arachico alla soluzione di colorante, cosa che consentiva anche di variare secondo le necessità la concentrazione del colorante nello strato, e quindi l'assortimento della radiazione incidente. Gli strati monomolecolari così preparati non presentavano una diffusione apprezzabile del colorante, anche dopo varie settimane, probabilmente per il buon ancoraggio offerto dalle lunghe catene idrocarburiche.

Consideriamo ora due tra i nostri argomenti di ricerca: le onde stazionarie che si formano davanti a uno specchio e l'onda che penetra nel mezzo riflettente nella situazione di « riflessione totale ». In entrambi i casi il detettore che misura l'intensità del campo elettrico associato alla radiazione è costituito da uno strato monomolecolare di colorante fluorescente. L'assorbimento della radiazione elettromagnetica incidente da parte del colorante è proporzionale al quadrato dell'intensità del campo elettrico e quindi anche la riemissione per fluorescenza dell'energia assorbita è proporzionale alla stessa quantità. L'assorbimento da parte del colorante è sufficientemente piccolo da non alterare il campo elettromagnetico che si sta studiando.

La tecnica usata dall'autore e dai suoi colleghi per depositare strati monomolecolari deriva dal procedimento inventato più di 30 anni fa da Katharine B. Blodgett e Irving Langmuir dei laboratori di ricerca della General Electric. Le fasi successive sono rappresentate schematicamente in prospettiva (*sopra*) e in sezione trasversale molto ingrandita (*sotto*). Dapprima si sparge sulla superficie dell'acqua una soluzione di un tensioattivo, quale un acido grasso a catena lineare lunga (*1*). Le molecole del tensioattivo insolubile si dispongono spontaneamente in strato monomolecolare con i gruppi idrofili rivolti verso l'acqua. Applicando una piccola forza alla barriera galleggiante (*2*) la molecole di acido grasso vengono spinte l'una contro l'altra, formando uno strato monomolecolare continuo. Poi si immerge una lastra di vetro attraverso la superficie ricoperta dal film (*3*), che non aderisce alla superficie idrofila del vetro. Risollevando la lastra (*4*) i gruppi carbossilici idrofili dell'acido aderiscono al vetro e si ottiene la deposizione di uno strato monomolecolare. Riimmergendo la lastra, ora idrofila, si ha l'adesione di un secondo strato monomolecolare (*5*) e un terzo si deposita tornando a sollevarla (*6*). La cosa si può ripetere per molte volte, fino a ottenere la sovrapposizione di centinaia di strati monomolecolari. Notare che, man mano che il film aderisce alla lastra, il galleggiante si sposta in avanti spinto dal peso, assicurando la continuità dello strato monomolecolare.

Lo spessore di uno strato multiplo si misura con un metodo interferometrico. Il film, deposto su uno specchio semitrasparente è irraggiato con una radiazione monocromatica con fronte d'onda piano, che subisce una riflessione multipla tra lo specchio e l'interfaccia acido grasso-aria. Le onde trasmesse attraverso lo specchio si sommano in un'onda risultante, la cui ampiezza dipende appunto dallo spessore del multistrato. Le frecce rappresentano ognuna la direzione di propagazione di un'onda piana.

La trasmissione della radiazione attraverso un multistrato, come nella figura in alto, varia periodicamente con il numero degli strati che lo compongono. Il numero di strati che occorre depositare per passare da un massimo (o minimo) di trasmissione al successivo corrisponde alla metà della lunghezza d'onda della radiazione incidente. In questo esperimento è stata usata radiazione con lunghezza d'onda di 5461 angstrom.

La posizione del monostrato fluorescente può essere variata a piacere rispetto alla superficie riflettente. Cambiando il numero di strati di acido grasso sottostanti e quindi, misurando l'intensità della fluorescenza, è possibile studiare la distribuzione spaziale dell'intensità del campo elettrico e conseguentemente il comportamento della radiazione elettromagnetica. Dato che i monostrati di acido grasso sono molto sottili se paragonati alla lunghezza d'onda delle radiazioni visibili, la risoluzione con cui si misura la distribuzione spaziale dell'energia del campo è piú che sufficiente per i nostri fini.

Quando una radiazione incide su una superficie riflettente, le onde riflesse interferiscono con le onde incidenti e si forma una cosiddetta onda stazionaria, analoga al movimento di una corda vibrante con un estremo fisso (*si veda la figura a pagina 88*). A determinate distanze dalla superficie riflettente i campi delle onde incidente e riflessa si annullano formando i « nodi » dell'onda stazionaria. A mezza strada tra i nodi, i campi si trovano invece « in fase » e si sommano negli internodi. Pertanto la distanza tra due nodi (o tra due internodi) equivale a metà della lunghezza d'onda della radiazione e, nel caso delle radiazioni visibili, è di qualche centinaio di angstrom. L'esistenza di un'onda stazionaria di fronte a una superficie riflettente era stata osservata già nel 1890 dal fisico tedesco Otto Wiener. Con la nostra tecnica è stato possibile studiarne accuratamente il comportamento.

Su uno specchio di elevato potere riflettente si depositano piú monostrati di acido grasso con la solita tecnica, facendo però in modo che ognuno di essi sia qualche millimetro piú corto del precedente. Si ottiene cosí una serie di gradini sui quali si deposita, come un tappeto, uno strato di colorante fluorescente.

Irraggiando la lastrina con un fascio monocromatico molto intenso, ottenibile usando una lampada a mercurio ad alta pressione, di fronte allo specchio si forma un'onda stazionaria (*si veda la figura a pagina 81*): alcuni dei gradini del multistrato verranno a trovarsi in corrispondenza con i nodi dell'onda stazionaria e quindi non si avrà fluorescenza del colorante, altri invece si troveranno in corrispondenza dei massimi di intensità del campo elettrico e qui si avrà un massimo nell'intensità della fluorescenza.

La fluorescenza dei gradini intermedi varierà in funzione dell'andamento del campo dell'onda stazionaria. I vari colori che si osservano in corrispondenza dei diversi gradini sono dovuti all'influenza dello specchio sulla fluorescenza, come si spiegherà piú avanti. Le nette differenze nella fluorescenza dei vari gradini provano che lo strato fluorescente è rimasto al suo posto dopo la deposizione. Se il colorante si fosse diffuso entro gli strati sottostanti la fluorescenza risulterebbe uniforme.

Le onde stazionarie che si formano davanti a uno specchio sono state usate recentemente per lo studio di brevissimi impulsi laser (si veda l'articolo « *Luce laser* » di A. L. Schawlow, in « Le Scienze », n. 4, dicembre 1968). La nostra tecnica potrebbe forse riuscire utile per approfondire lo studio di questi impulsi ultrabrevi.

Consideriamo ora il secondo fenomeno studiato: l'onda che si forma all'interfaccia tra due mezzi trasparenti in caso di riflessione totale.

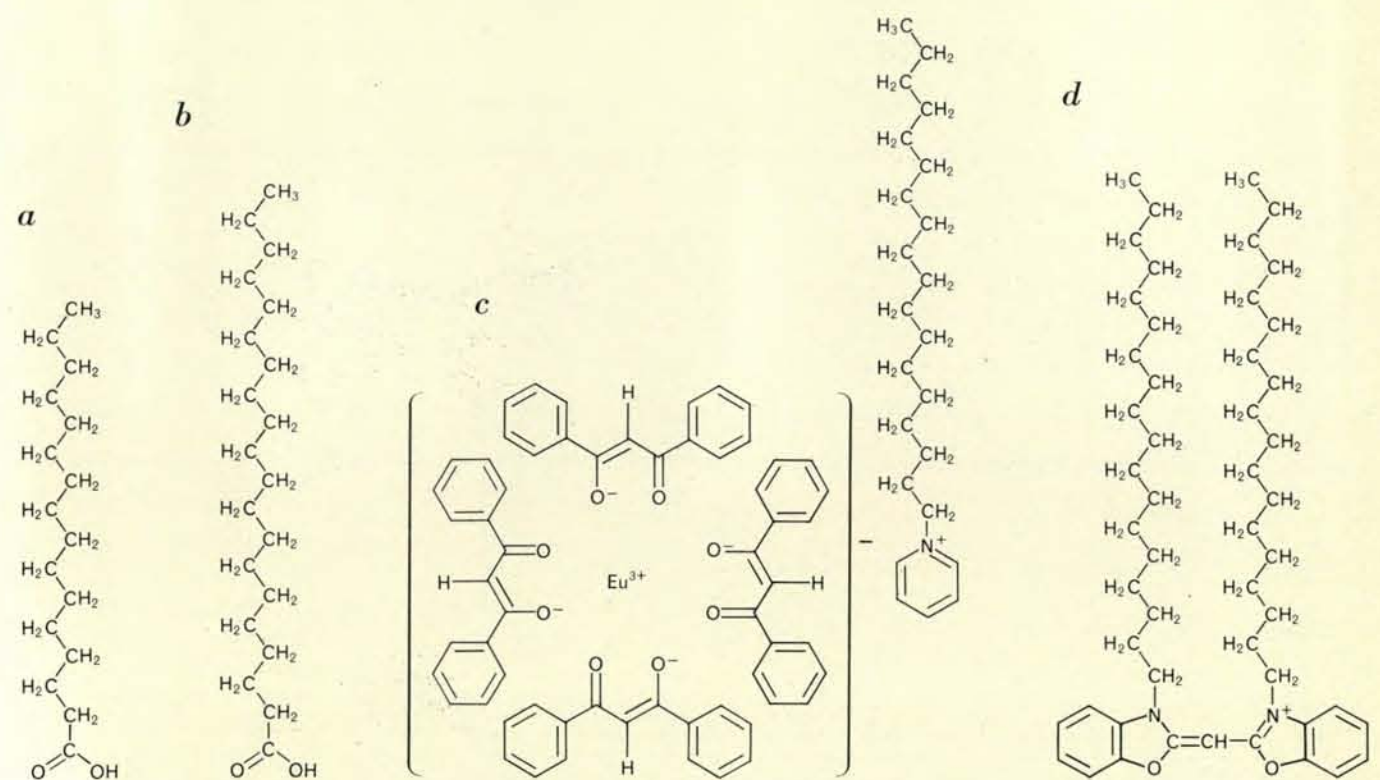
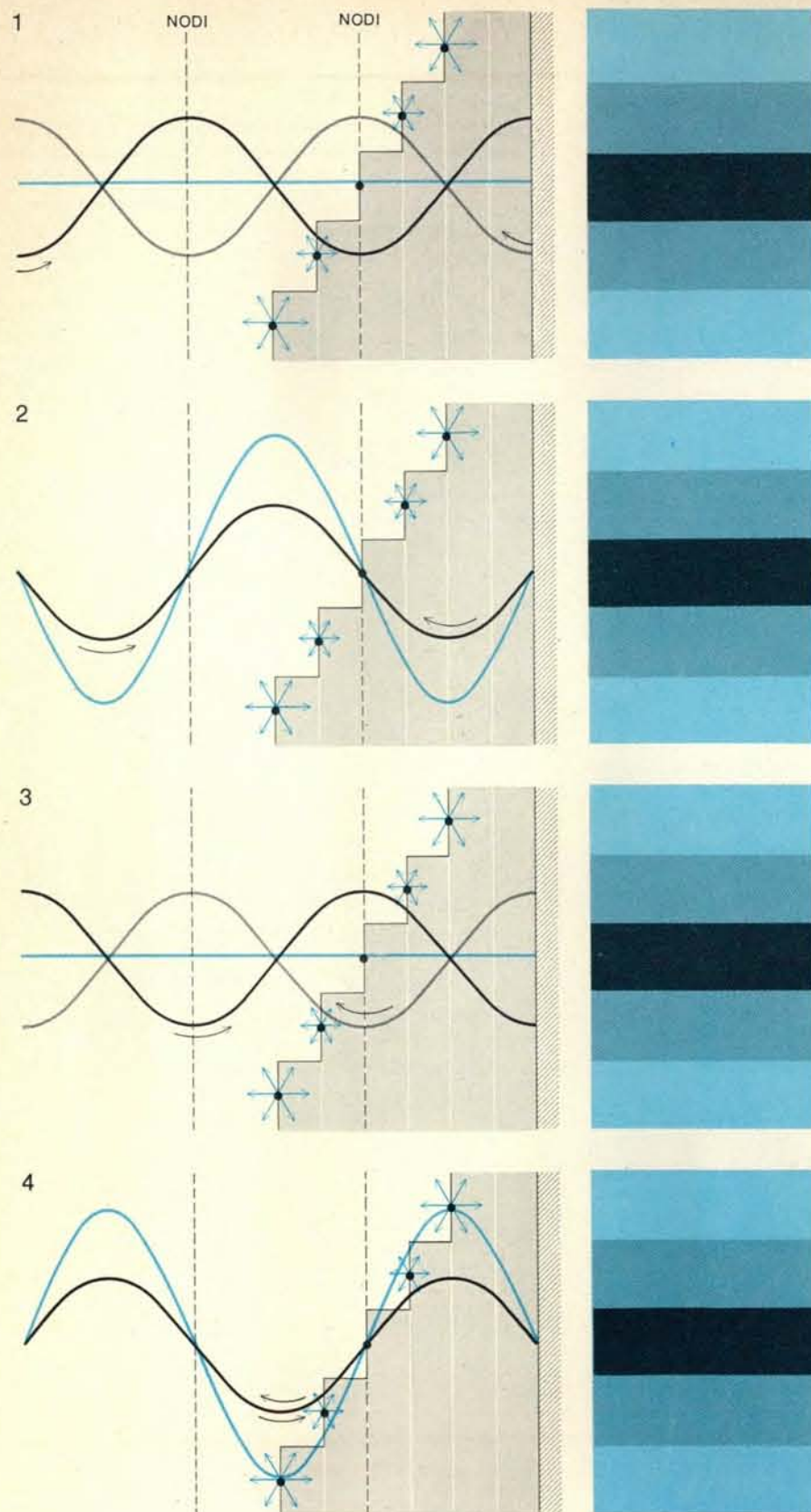
Quando una radiazione visibile incide sull'interfaccia tra due mezzi trasparenti, ma di diverso indice di rifrazione, con un angolo superiore a un certo valore critico, essa viene totalmente riflessa, come su uno specchio. Ma anche in questa situazione si osserva un'onda che si propaga oltre l'interfaccia.

Questo fenomeno ha attirato l'attenzione di molti ottici, a partire da Isaac Newton, che per primo ne verificò l'esistenza. Ma finora la propagazione di quest'onda non era mai stata dimostrata con esperienze non ambigue e valide sul piano quantitativo in cui si impiegasse una radiazione visibile. Secondo le teorie della radiazione, quest'onda diminuisce esponenzialmente di intensità al crescere della distanza dell'interfaccia e praticamente scompare nello spazio di poche lunghezze d'onda. Ma questa distanza pur breve è sufficientemente grande in confronto allo spessore di un monostrato di acido grasso e quindi la nostra tecnica consente lo studio del fenomeno con buona risoluzione.

Per far questo, abbiamo depositato uno strato monomolecolare di colorante fluorescente su una lastra di vetro, ricoprendolo successivamente con un numero variabile di monostrati di CdC_{20}, il cui indice di rifrazione è praticamente identico a quello del vetro, cosicché la radiazione usata non viene deviata all'interfaccia tra questi due mezzi. Il tutto veniva immerso in un liquido a elevato indice di rifrazione.

Si faceva poi incidere sul multistrato, attraverso il liquido, un fascio di radiazione monocromatica con lunghezza d'onda di 4050 angstrom (colore viola), a un angolo superiore a quello di riflessione totale (*si veda la figura a pagina 89 in alto*). Il multistrato funge sia da mezzo a basso indice di rifrazione che da distanziatore per il monostrato fluorescente. Si può cosí studiare l'intensità del campo in funzione della distanza dell'interfaccia, misurando la fluorescenza gialla del colorante. Si è potuto determinare sperimentalmente che l'intensità della radiazione diminuisce esponenzialmente al crescere della distanza dell'interfaccia, in accordo con la teoria. Si è potuta verificare anche la relazione esistente tra l'assorbimento della radiazione e l'angolo di incidenza, che era stata prevista dalla teoria.

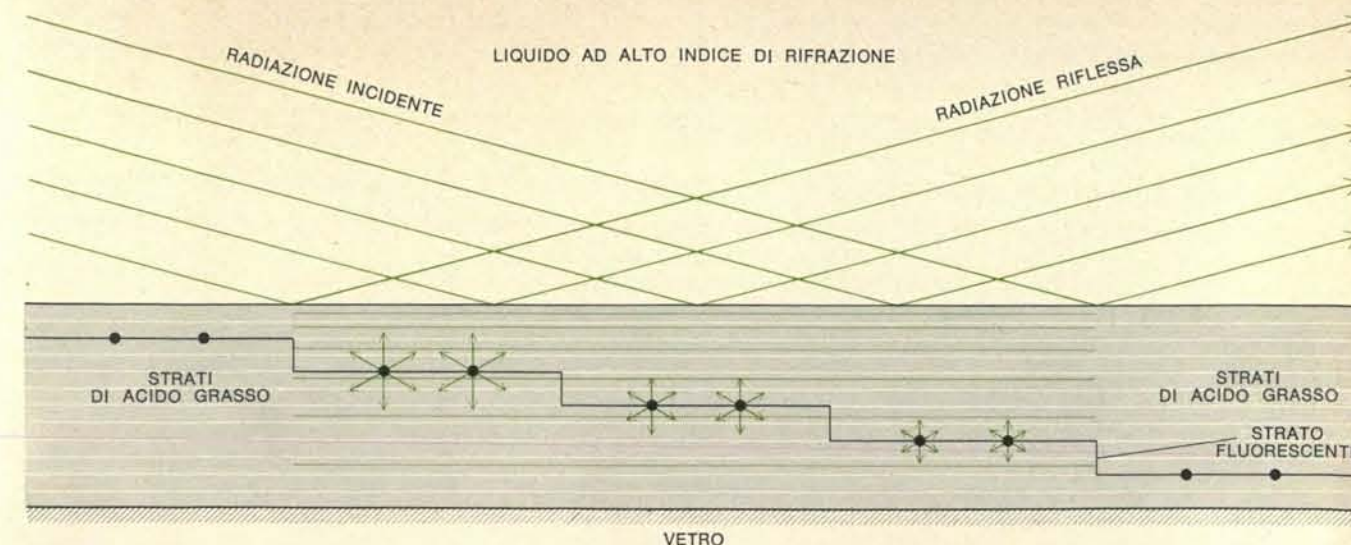
Fin qui si è descritto l'impiego dei coloranti fluorescenti come detettori dell'intensità della radiazione eccitante. Ora, invece, prendiamo in considerazione il processo di emissione della radiazione di fluorescenza. L'emissione delle onde radio da parte di antenne è descritta perfino in testi elementari e, dato che anch'esse sono radiazioni elettromagnetiche è logico chiedersi se vi sia qualche analogia tra antenne radio e molecole che irraggiano. La risposta è affermativa: v'è una stretta parentela tra i due fenomeni.

Le antenne radio emettono una radiazione coerente, cioè le onde emesse nelle varie direzioni hanno un preciso rapporto di fase, rivelato da fenomeni di interferenza. Consideriamo ora la ra-

Struttura di alcune molecole adatte per la formazione di strati monomolecolari. L'acido palmitico (*a*) e l'acido arachico (*b*) appartengono alla classe degli acidi grassi a lunga catena lineare. Il complesso di europio (*c*) e le cianine a lunga catena (*d*) sono stati sintetizzati appositamente per gli esperimenti in questione. Queste ultime tre sostanze sono fluorescenti.

Quando una radiazione incide su un riflettore si forma un'onda stazionaria. Il campo elettrico risultante (*in colore*) è somma dei campi della radiazione incidente (*in nero*) e di quella riflessa (*in grigio*). A determinate distanze dallo specchio il campo si annulla e le molecole fluorescenti che si trovano in quei punti, i «nodi», non vengono eccitate. Le molecole tra un nodo e l'altro vengono invece eccitate in ragione del quadrato dell'ampiezza del campo elettrico esistente nel punto in cui si trovano. Ciò spiega le differenze dell'intensità della fluorescenza dallo strato deposto a gradini della figura a pagina 81. La distanza tra due nodi equivale a mezza lunghezza d'onda.

diazione emessa da una molecola eccitata posta di fronte a uno specchio (*si veda la figura a pagina 89 in basso*). L'onda emessa interferisce con la parte della stessa onda riflessa dallo specchio. L'interferenza può essere costruttiva o distruttiva, in funzione del diverso cammino percorso, che a sua volta dipende dall'angolo di incidenza sullo specchio. Di conseguenza l'intensità dell'emissione è grande in certe direzioni, mentre in altre, in cui le radiazioni diretta e riflessa sono in opposizione di fase, è invece nulla. Ne segue che l'intensità della fluorescenza di una molecola posta di fronte a uno specchio deve avere una netta dipendenza dall'angolo di rilevamento.

Abbiamo studiato questo fenomeno utilizzando monostrati di un complesso dell'europio, un elemento del gruppo delle terre rare. In questo complesso i quattro leganti organici che circondano l'atomo del metallo assorbono radiazione ultravioletta a 3666 angstrom e trasferiscono rapidamente l'energia assorbita allo ione europio, che viene eccitato e si riporta al livello energetico di base emettendo una radiazione avente lunghezza d'onda di 6120 angstrom e dando quindi origine a una fluorescenza rossa.

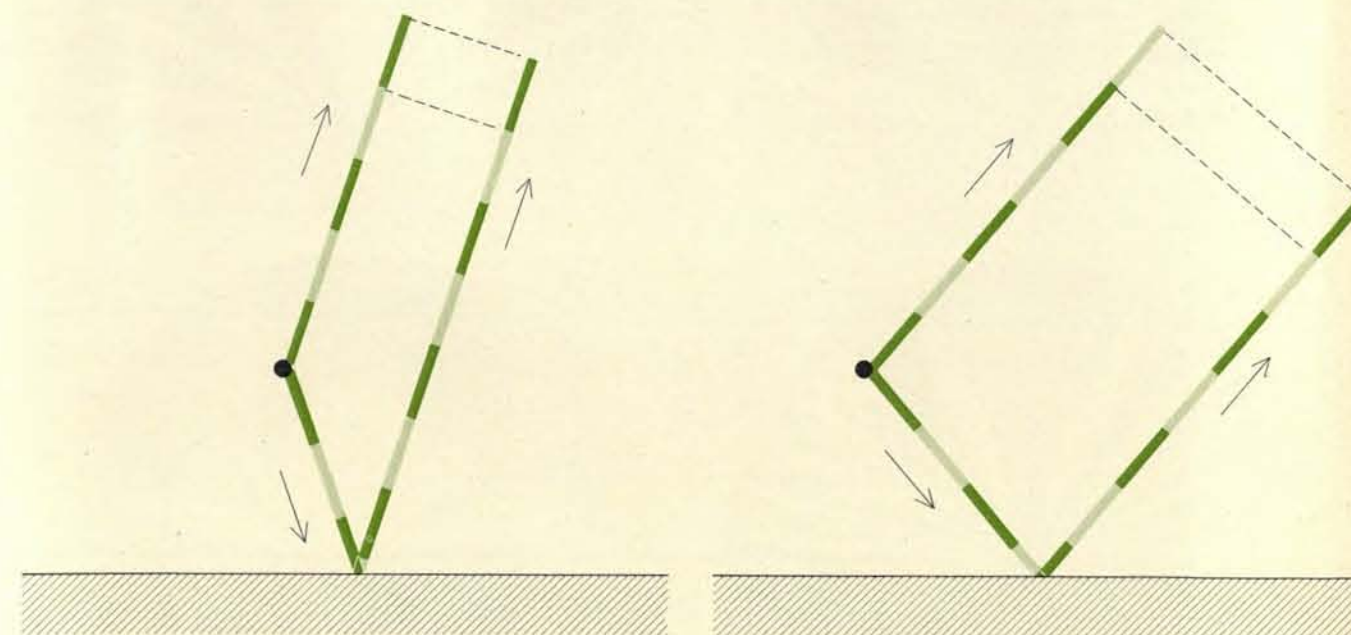
Nei nostri esperimenti il monostrato fluorescente veniva deposto sopra un multistrato di acido grasso che fungeva da distanziatore di spessore noto da una lastrina ricoperta di uno strato riflettente d'oro o d'argento. Per evitare la rifrazione delle onde sull'interfaccia aria-multistrato, il tutto veniva immerso in un liquido con lo stesso indice di rifrazione dell'acido grasso contenuto in un recipiente cilindrico. La lastra veniva quindi eccitata con radiazione ultravioletta e si misurava la distribuzione angolare della fluorescenza, mantenendo costante la direzione del fascio di eccitazione. Abbiamo notato che i grafici di distribuzione angolare della fluorescenza sono molti simili a quelli dell'intensità di emissione di un'antenna radio posta di fronte a un riflettore (*si veda la figura a pagina 90*).

Vi sono diversi tipi di antenne radio. Quello piú semplice è costituito da un dipolo elettrico oscillante. Tipi piú complicati sono costituiti da un dipolo magnetico o da un quadrupolo elettrico e ognuno di questi tipi di antenna possiede una distribuzione angolare caratteristica dell'intensità di emissione. È interessante notare che i vari tipi di molecole fluorescenti si comportano in maniera del tutto simile, cosa che ci ha permesso di determinare la

L'onda trasmessa che accompagna la riflessione totale viene completamente assorbita a una distanza di poche lunghezze d'onda dall'interfaccia su cui avviene la riflessione. La sua intensità si misura facilmente con uno strato monomolecolare fluorescente.

natura dell'oscillatore molecolare mediante misure di distribuzione angolare dell'intensità della fluorescenza. Esperimenti simili erano già stati fatti nel 1911 dal fisico ungherese Paul Selenyi, ma con la nostra tecnica è stato possibile effettuare misure molto precise.

Noi abbiamo calcolato la distribuzione angolare della fluorescenza rossa dell'europio, nell'ipotesi che la sorgente di radiazione fosse costituita da dipoli elettrici orientati casualmente. La coincidenza della distribuzione calcolata, con scarti di poche unità per cento, con quella effettivamente misurata prova che la fluorescenza è dovuta effettivamente a un dipolo elettrico.

La fluorescenza rossa del complesso è accompagnata da una debole fluorescenza arancione (lunghezza d'onda di 5920 angstrom), che si può studiare eliminando la prima con un filtro interferenziale. La distribuzione angolare di questa fluorescenza è completamente diversa da quella trovata in precedenza e dimostra che questa radiazione è dovuta all'oscillazione di un dipolo magnetico, in accordo con quanto trovato seguendo altri metodi.

La distribuzione angolare dell'emissione di un oscillatore posto di fronte a un riflettore dipende dalla lunghezza d'onda della radiazione emessa. Ciò si può controllare impiegando sostanze fluorescenti che emettano con continuità in tutto lo spettro. Simili sostanze, poste di fronte a uno specchio emettono, in una data direzione, radiazioni la cui lunghezza d'onda dipende dalla distanza dallo specchio. Ciò spiega la diversa colorazione dei gradini nella figura a pagina 81. Cambiando l'angolo d'osservazione, cambiano pure i colori che si osservano, in quanto varia la distanza tra la molecola emettitrice e il punto di riflessione e quindi anche la lunghezza delle onde per cui si ha interferenza costruttiva o distruttiva.

L'interferenza tra l'onda irradiata da un'antenna (o da una molecola eccitata), posta di fronte a un riflettore, e la parte dell'onda riflessa produce una caratteristica distribuzione angolare dell'intensità del campo risultante. A seconda della differenza tra i percorsi dell'onda diretta e riflessa, l'interferenza può essere costruttiva (*a sinistra*) o distruttiva (*a destra*).

Il riflettore non si limita a influenzare la distribuzione angolare della radiazione emessa dall'antenna postagli davanti: ha influenza pure sulla probabilità dell'emissione. Per spiegarci questo fenomeno, consideriamo un'antenna al centro di un riflettore semisferico (*si veda la figura in alto a pagina 91*). Se il raggio di curvatura è tale che le onde diretta e riflessa siano in fase, le loro ampiezze si sommano e l'intensità (che è proporzionale al quadrato dell'ampiezza) diviene 4 volte maggiore di quella che si avrebbe senza riflettore. Questo è vero per tutte le direzioni, quindi la potenza totale dell'emissione è il doppio di quella che si avrebbe senza riflettore. Se però il raggio della sfera viene variato di 1/4 di lunghezza d'onda, le onde diretta e riflessa si trovano in opposizione di fase e l'antenna non emette per nulla.

Trasferiamo queste considerazioni a una molecola eccitata. In questo caso è possibile misurare la durata media del livello eccitato, che è inversamente proporzionale alla probabilità che avvenga il suo decadimento e quindi l'emissione di radiazione. Nel primo dei due casi precedenti la vita media del livello viene dimezzata dalla presenza del riflettore, mentre nel secondo diventa infinita: la molecola rimane eccitata indefinitamente, non potendo irraggiare. In realtà esistono processi di disattivazione senza irraggiamento, in competizione con la fluorescenza, che non consentono la durata infinita dei livelli eccitati.

Per lo studio di questo fenomeno abbiamo usato uno specchio piano, anziché sferico, e abbiamo misurato la vita media dei livelli eccitati in funzione della distanza del monostrato fluorescente dallo specchio. La disposizione sperimentale era molto simile a quella usata per lo studio della distribuzione angolare della fluorescenza. Il complesso di europio si è dimostrato particolarmente adatto per questo studio, in quanto la vita media del livello eccitato è circa 1/1000 di secondo, durata facilmente misurabile con precisione.

I risultati confermano la prevista dipendenza della vita media dalla distanza molecola-riflettore (*si veda la figura in basso a pagina 91*). Dato che lo specchio influenza solo la probabilità dell'emissione di radiazione e non i processi di disattivazione competitivi, è possibile ottenere informazioni sulla probabilità relativa dei due tipi di decadimento. Ci si deve attendere che la dipendenza della vita media del livello eccitato dalla distanza molecola-riflettore sia massima in assenza di decadimenti non radiativi e diventi tanto minore quanto piú aumenta la probabilità di tali processi. Si può cosí determinare il « rendimento quantico » dello stato eccitato, altrimenti difficilmente misurabile. Nel caso del complesso di europio esso è del 70 per cento.

L'influenza del riflettore sull'emissione di radiazione è indipendente dal tipo delle onde irraggiate. Quindi se possedessimo uno specchio in grado di riflettere i raggi gamma sarebbe possibile variare il tempo di dimezzamento dei nuclei radioattivi!

Un altro aspetto interessante di questi esperimenti deriva dal fatto che una molecola eccitata può emettere un solo quanto di radiazione (fotone). Dato che nelle nostre condizioni sperimentali solo un piccola frazione delle molecole di colorante era eccitata, esse emettevano il loro fotone indipendentemente l'una dall'altra. Eppure i fenomeni di interferenza avevano esattamente l'entità prevista, evidentemente non richiedono l'emissione contemporanea di due fotoni.

Ciò sta a dimostrare che anche un solo quanto di radiazione è in grado di produrre fenomeni di interferenza, per le sue proprietà ondulatorie.

Le distribuzioni angolari della fluorescenza di una molecola eccitata posta di fronte a uno specchio somigliano alle distribuzioni angolari di un'antenna radio. Le distribuzioni *a*, *b* e *c* sono dovute alla fluorescenza rossa di uno strato monomolecolare del complesso di europio a varie distanze dallo specchio, corrispondenti rispettivamente a 142, 256 e 446 strati di acido grasso. Questa fluorescenza è dovuta all'oscillazione di un dipolo elettrico. La distribuzione *d* è stata ottenuta come la *a*, ma misurando l'intensità della debole fluorescenza arancione, che accompagna la fluorescenza rossa del complesso. Quest'ultima distribuzione è diversa dalla prima, in quanto la fluorescenza arancione è dovuta all'oscillazione di un dipolo magnetico.

L'influenza dello specchio sulla fluorescenza di una molecola è un effetto a lungo raggio, nel senso che è ancora misurabile per distanze molecola-riflettore pari a diverse lunghezza d'onda. Per questa ragione è possibile spiegare il fenomeno ragionando solo sul campo elettromagnetico irradiato. Infatti un dipolo elettrico oscillante ha associato anche un campo la cui intensità diminuisce con il cubo della distanza dall'oscillatore, e quindi si annulla rapidamente. Ma nei pressi del dipolo questo campo è molto piú intenso di quello che si propaga a lunga distanza.

Ci si deve attendere di trovare questo campo a breve raggio d'azione anche nei pressi di una molecola eccitata e di poterne misurare l'intensità con l'impiego di sostanze fluorescenti. A questo scopo abbiamo depositato su un vetro un monostrato di una cianina con fluorescenza blu. Si è poi depositato un secondo monostrato di un'altra cianina, separato dal primo da pochi strati di acido grasso. La seconda cianina è stata scelta in modo da poter essere eccitata proprio dalla radiazione emessa dalla prima e presenta una fluorescenza gialla. Il secondo strato può quindi servire come detettore del campo a corto raggio nei pressi del primo strato (*si veda la figura a pagina 82*). I dati ottenuti dimostrano l'esistenza del campo previsto nei pressi della molecola eccitata e confermano che la sua intensità diminuisce in maniera inversamente proporzionale al cubo della distanza.

In questo esperimento l'energia di eccitazione viene trasferita dalle molecole a fluorescenza blu (dette sensibilizzatori) a quelle a fluorescenza gialla (dette accettori). È notevole che il semplice meccanismo di interazione col campo a corto raggio sia riuscito a dimostrare, anche da un punto di vista quantitativo, quel processo di trasporto di energia detto correntemente « trasferimento non radiativo per risonanza ».

Con uno specchio sferico è possibile dimostrare l'influenza del riflettore sull'irraggiamento di un'antenna. Se il raggio della sfera è tale che l'onda diretta e riflessa siano in fase (*a sinistra*) si ha interferenza costruttiva e l'intensità dell'emissione risulta quattro volte maggiore di quella della sola onda diretta. Poiché questo vale in tutte le direzioni di fronte allo specchio, ne deriva che la probabilità di una emissione è raddoppiata per la presenza del riflettore. Se invece le dimensioni del riflettore sono tali che le due onde si vengano a trovare in opposizione di fase (*a destra*) non v'è piú irraggiamento e la probabilità di una emissione cade a zero. Dato che la durata media di un livello eccitato è inversamente proporzionale alla probabilità di emissione, la vita media viene dimezzata nel primo caso, mentre diviene infinita nel secondo.

I complessi di piú strati monomolecolari descritti in questo articolo possono essere considerati dei cristalli artificiali, la cui composizione può essere variata a piacere entro limiti piuttosto ampi. Il metodo di deposizione in strati successivi consente di realizzare un ordine prefissato nella direzione di accrescimento, cosa difficile da ottenersi in altro modo. Quindi è ragionevole prevedere applicazioni dei multistrati in moltissimi campi di ricerca.

Multistrati prodotti da soluzioni contenenti ioni come il bario o il piombo presentano una spaziatura di circa 50 angstrom tra i piani contenenti gli atomi pesanti. Sono stati utilizzati in analizzatori per raggi X molli e possono servire per futuri laser a raggi X. Strutture stratificate di molecole organiche hanno anche grande importanza biologica ed è possibile preparare membrane artificiali a vari strati di fosfolipidi e proteine, il cui studio è utilissimo per la comprensione del funzionamento delle membrane biologiche.

La vita media dei livelli eccitati nel complesso di europio dipende dalla distanza delle molecole da uno specchio piano, che è stata variata con l'interposizione di strati monomolecolari di acido grasso, in salti di 52,8 angstrom (due strati monomolecolari). La curva rappresenta la dipendenza prevista teoricamente, i cerchi i valori sperimentali.

Il "secondo suono" nell'elio solido

Si riteneva che onde termiche analoghe alle onde acustiche ordinarie ("primo suono") esistessero solamente nell'elio liquido. Ora sono state osservate anche nell'elio solido e ciò può accadere in altre sostanze

di Bernard Bertman e David J. Sandiford

Sin dal XIX secolo fisici e ingegneri hanno pensato al flusso di calore nei materiali come a un processo di diffusione. Il flusso di calore è allora proporzionale alla differenza di temperatura per unità di lunghezza del materiale che esiste tra la sorgente e il pozzo di calore; la costante di proporzionalità dipende dal materiale ed è chiamata conducibilità termica. Recenti esperimenti hanno però mostrato che il calore può fluire qualche volta in un modo completamente diverso. Ricercatori della Duke University e del National Laboratory of Los Alamos hanno riferito che, in opportune condizioni, il calore può fluire nell'elio solido in forma di onda. (L'elio solido è ottenuto comprimendo elio liquefatto con metodi criogenici). Del resto già qualche anno prima era stato osservato che nell'elio liquido « superfluido » il flusso di calore non era un processo di diffusione e questo comportamento termico ondulatorio era stato chiamato « secondo suono ». È opinione recente che il fenomeno del secondo suono sia molto piú generale di quanto supposto e che, in condizioni opportune, il secondo suono stesso sia osservabile in altri materiali diversi dall'elio liquido e solido.

Come si può visualizzare la differenza tra il calore che fluisce come un'onda e il calore che fluisce per diffusione? È utile a questo scopo considerare il seguente esperimento ipotetico. Un lungo bastone, con un'estremità in un secchio di ghiaccio, è mantenuto alla temperatura uniforme di zero gradi centigradi (*si veda la figura in basso a pag. 94*). Si riscalda per una frazione di secondo l'estremità superiore del bastone con una sorgente di calore (per esempio una fiamma di butano). Come si comporta la temperatura del bastone?

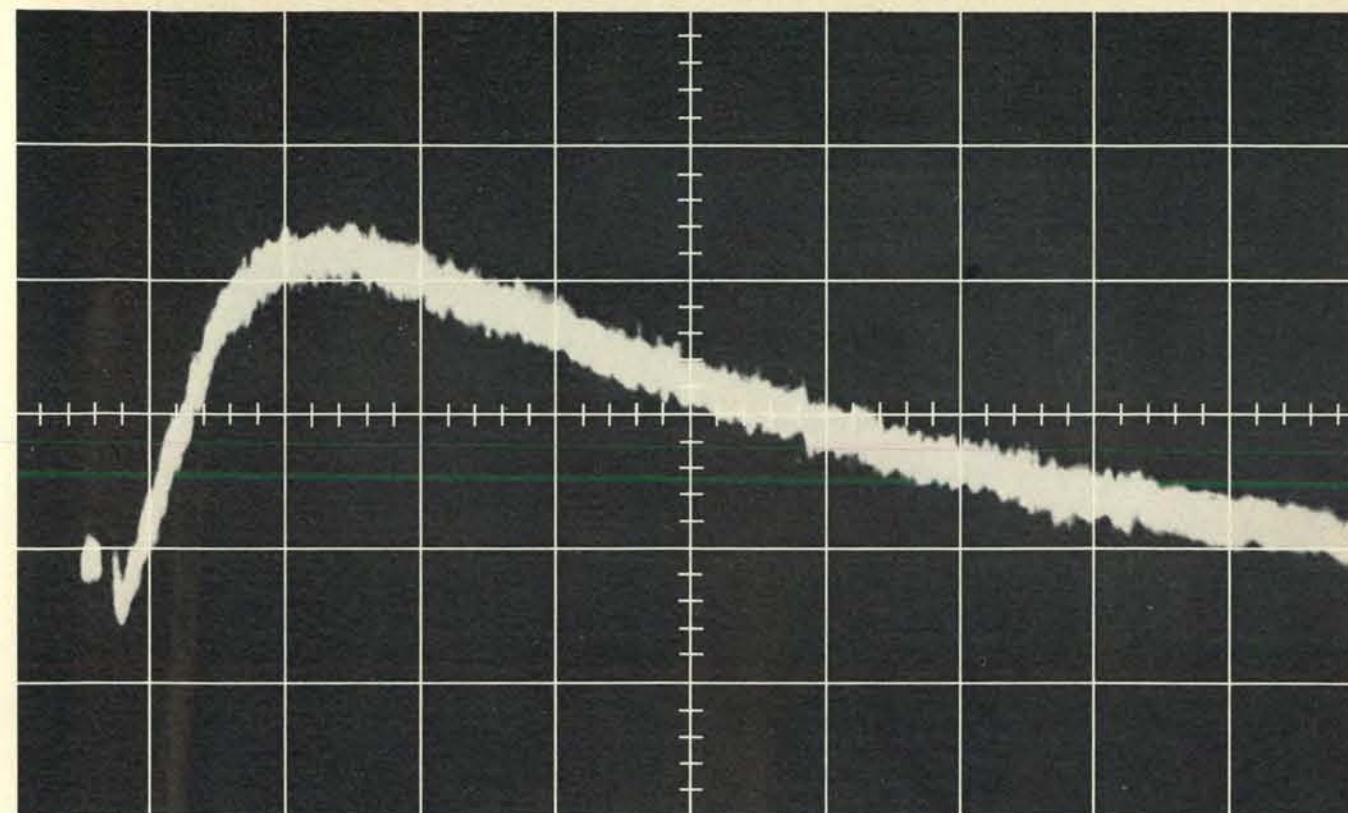
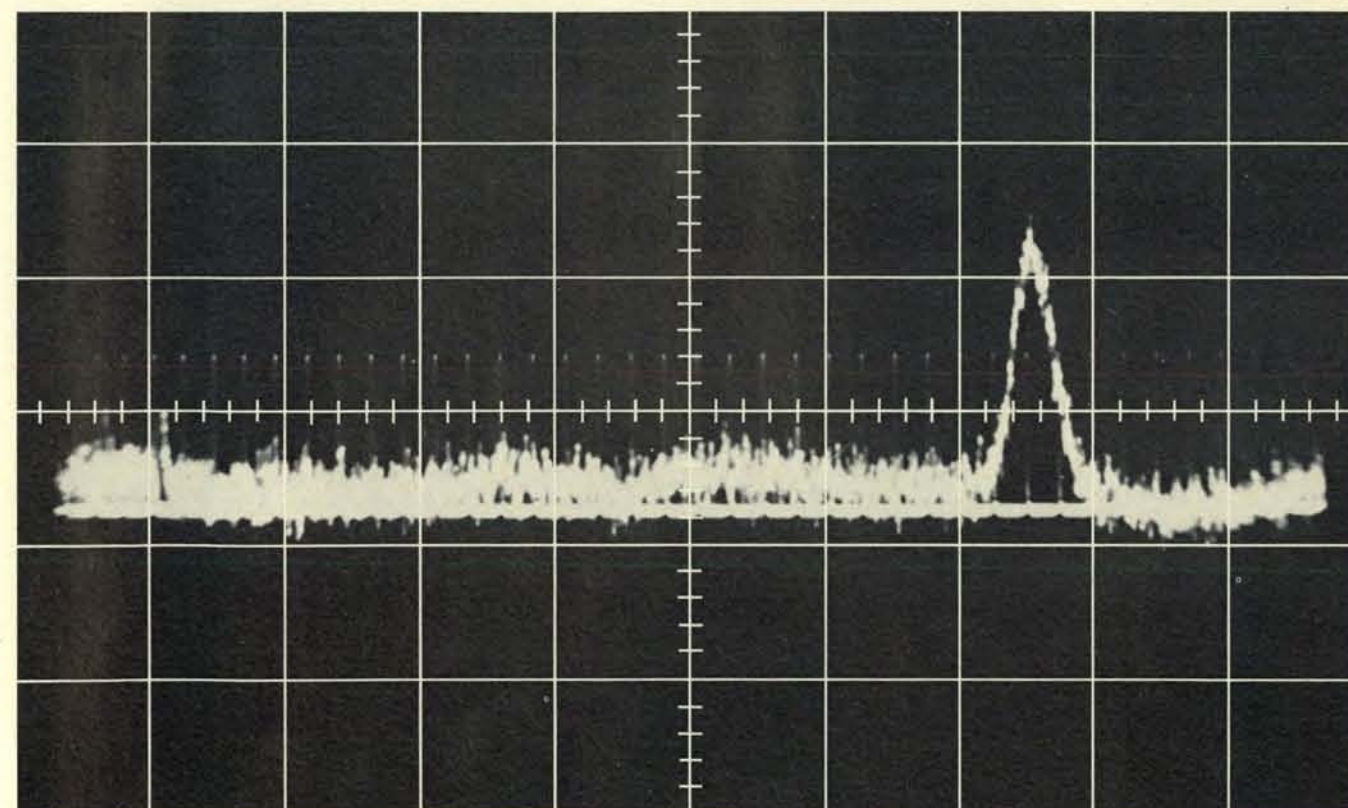
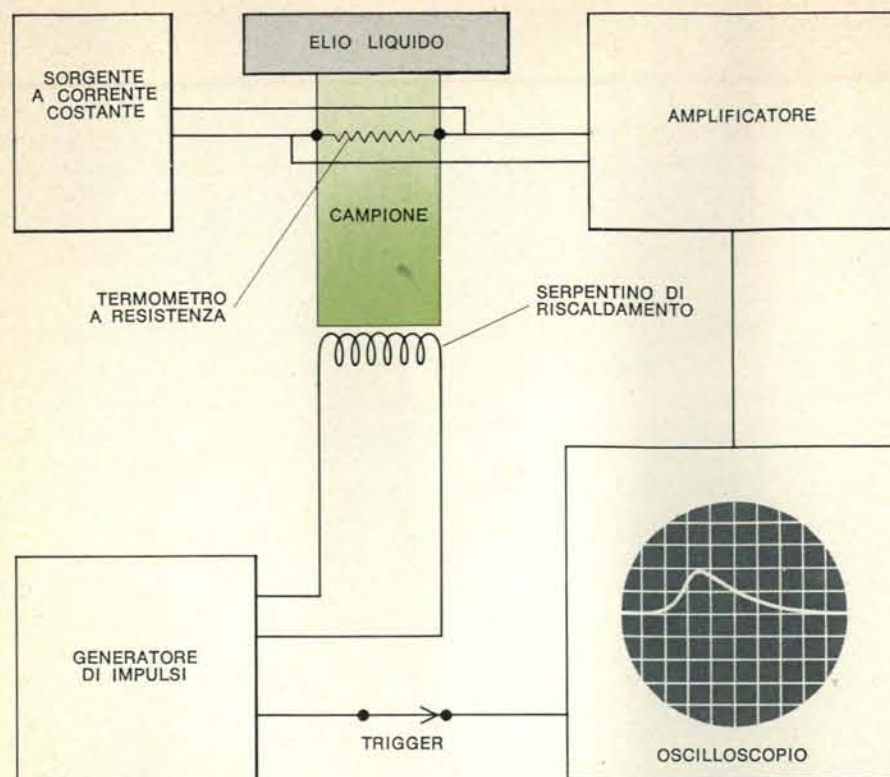
Nell'istante in cui l'estremità è riscaldata, questa parte del bastone subisce un aumento considerevole della temperatura; altrove la temperatura è ancora a zero gradi centigradi (*si veda la figura in alto a pag. 95*). Un attimo dopo l'energia termica aggiunta per mezzo della fiamma comincerà a fluire verso l'estremità posta nel ghiaccio, e il grafico della temperatura comincerà ad appiattirsi; ciò continuerà nei susseguenti intervalli di tempo via via che il calore fluisce dall'estremità piú calda del bastone verso quella piú fredda. Questo processo è inizialmente rapido. poiché il gradiente della temperatura è grande, ma rallenta gradualmente al diminuire della temperatura. Infine tutto il calore aggiunto lascia il materiale e il bastone ritorna alla temperatura di zero gradi centigradi.

Quale temperatura misurerà un termometro posto in una posizione fissa sul bastone, dopo che il calore è stato applicato all'estremità? Ovviamente si può ottenere questa informazione dal grafico della temperatura già descritto. L'ampiezza e la forma dell'impulso della temperatura dipenderanno dalla posizione del termometro; vicino all'estremità riscaldata l'impulso sarà piú alto e meno esteso nel tempo di quanto non accadrebbe vicino all'estremità piú fredda, dove l'aumento di temperatura è piú basso. Questi risultati sono caratteristici del flusso di calore per diffusione.

Se supponiamo invece che il calore possa muoversi attraverso il materiale sotto forma di onda, il grafico della temperatura apparirà completamente diverso (*si veda la figura in basso a pag. 95*). L'aumento di temperatura si muoverà con una velocità costante attraverso il materiale e tutto il calore lascerà immediatamente l'estremità calda del bastone. Il calore sarà allora contenuto in una piccola regione dello spazio che cambia la sua posizione in un modo molto semplice e la temperatura misurata in un qualsiasi punto fisso aumenterà e diminuirà bruscamente. Inoltre l'impulso di temperatura non varierà in altezza e spessore cambiando la posizione del termometro: solamente l'istante di arrivo sarà diverso. Sino a pochi anni or sono era noto un solo esempio di materiale in cui il calore si propagasse come un'onda: l'elio liquido superfluido a una temperatura vicina allo zero assoluto.

Espressioni come « flusso di calore » e « temperatura » hanno un significato approssimativo nella vita di ogni giorno. Per proseguire nella discussione sarà quindi utile avere una idea piú precisa di cosa significano queste espressioni per un fisico. Per una comprensione delle definizioni microscopiche e molecolari di queste grandezze consideriamo prima di tutto un gas. In un gas, come l'aria attorno a noi, ci sono 3×10^{19} molecole di azoto e ossigeno al centimetro cubo che si muovono in tutte le direzioni e si urtano l'una con l'altra. La distanza media tra le molecole nell'aria è circa 30 angstrom (1 angstrom vale 10^{-8} centimetri) e una molecola media, nell'aria a pressione e temperatura ambiente, tra due collisioni successive si muove in linea retta con una velocità di circa 300 metri al secondo. Lo spazio percorso in media da una molecola tra un urto e l'altro è di circa 1000 angstrom; questa distanza è chiamata cammino libero medio della molecola nel gas.

Dopo l'urto le due molecole che hanno partecipato alla collisione avranno velocità diverse in direzione e intensità, ma le nuove velocità devono essere tali da rispettare le leggi di conservazione della meccanica: l'energia totale delle molecole prima dell'urto deve essere eguale all'energia totale delle molecole dopo l'urto, e lo stesso deve accadere per la quantità di moto (impulso). Il movimento delle molecole è in ogni istante governato da queste leggi.

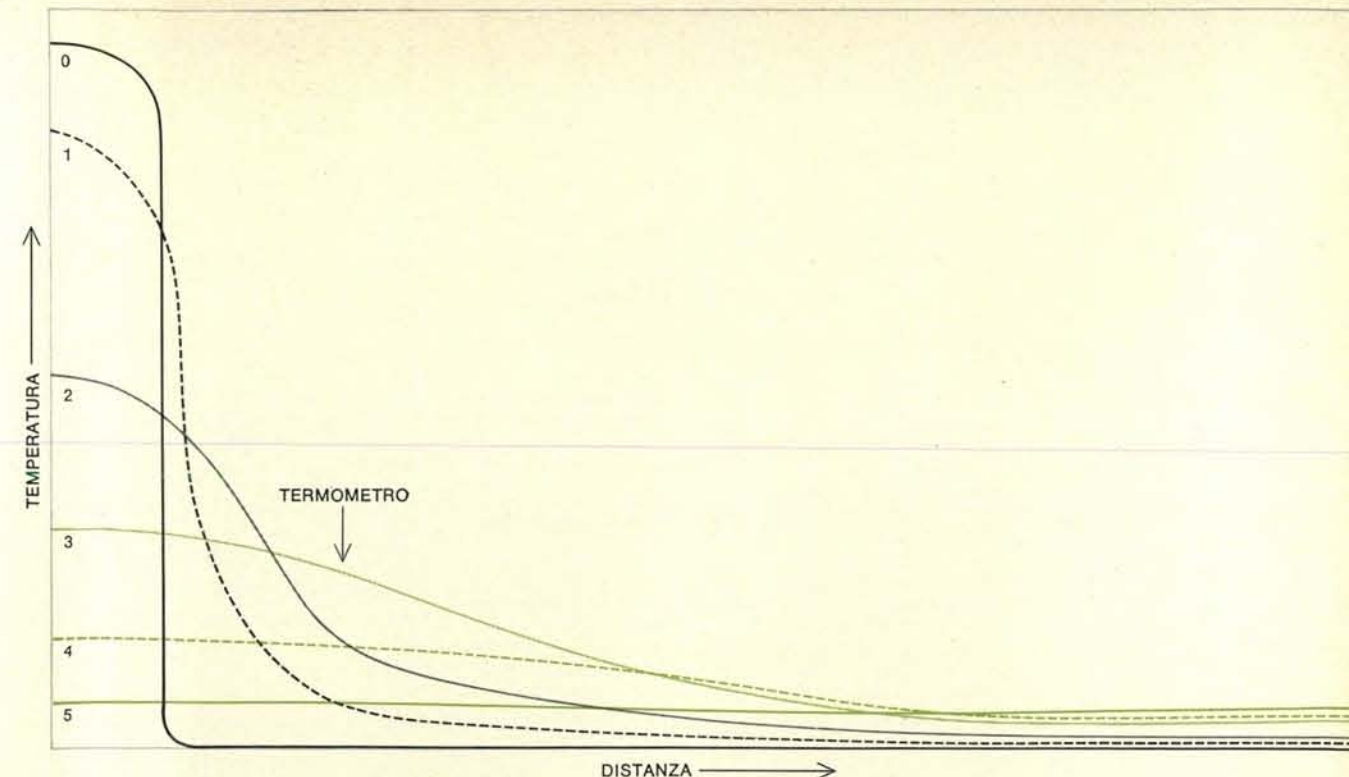
In un gas la maggior parte dell'energia è energia cinetica, sia di traslazione che di rotazione (se la molecola non ha simmetria sferica). La cinetica di un

Il flusso di calore per diffusione è evidenziato da questa traccia sull'oscilloscopio, che rappresenta la temperatura del materiale campione in funzione del tempo, misurata a una distanza fissata dall'estremità riscaldata. L'oscilloscopio è pilotato dall'impulso stesso di calore. Il graduale aumento e la caduta della temperatura al tempo di misura sono tipici della diffusione.

Il carattere ondulatorio del flusso di calore è rappresentato da questa traccia sull'oscilloscopio; il campione è costituito dall'elio liquido superfluido a circa 1 °K. La temperatura misurata in un punto fisso aumenta e diminuisce bruscamente in un certo istante dopo che l'impulso di calore è stato applicato all'estremità del campione. L'impulso di temperatura si muove con velocità costante attraverso il materiale e non varia in altezza cambiando la posizione del termometro; cambia solamente l'istante di arrivo del segnale. Questo comportamento è tipico delle onde termiche, che sono state chiamate originariamente onde del secondo suono per l'analogia diretta esistente con le onde del primo suono, cioè con le onde acustiche.

Apparato sperimentale usato per ottenere le tracce delle precedenti figure. Un generatore di impulsi elettrici provoca un impulso di calore sul campione e simultaneamente pilota l'oscilloscopio. La temperatura è misurata con un termometro a resistenza.

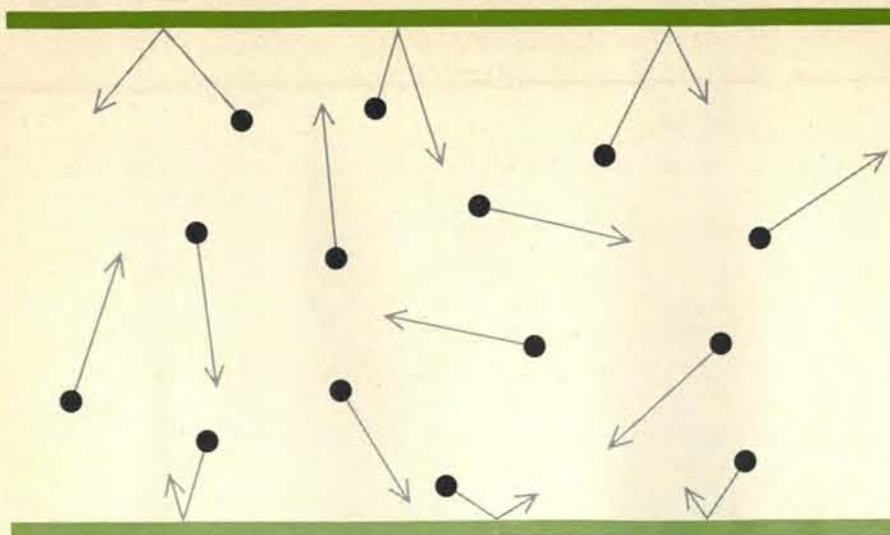
Il seguente esperimento è utile per la comprensione della differenza tra il calore che fluisce per diffusione e il calore che fluisce sotto forma di onda. Un lungo bastone con un'estremità in un secchiello di ghiaccio, è posto alla temperatura uniforme di zero gradi. L'altra estremità è riscaldata per una frazione di secondo con una fiamma di butano. La temperatura può essere misurata in due modi: lungo tutto il bastone in un istante fissato oppure in successivi intervalli di tempo in un punto fissato.

insieme di molecole (cioè di un gas) è del tutto particolare in quanto il gas stesso contiene molecole che si muovono casualmente in tutte le direzioni. Questo moto è chiamato agitazione termica microscopica, e deve essere distinto da un moto ordinato. Un esempio di moto ordinato si potrebbe avere nel caso che tutte le molecole possedessero la stessa velocità di 10 metri al secondo in direzione nord; questa situazione è rappresentata da un vento (un movimento dell'aria nel suo complesso) con una velocità di 10 metri al secondo in direzione nord; esso non altera però gli urti tra gli atomi, poiché le loro velocità relative rimangono inalterate. L'energia totale interna di un gas è data dalla somma di tutte le energie cinetiche casuali di ogni singola molecola.

Un importante risultato della fisica classica è la proporzionalità tra l'energia interna di un gas e la temperatura assoluta in gradi Kelvin (gradi centigradi sopra lo zero assoluto). Quindi se la temperatura dell'aria aumenta da 300 °K a 330 °K, l'energia cinetica del moto delle molecole aumenta del 10%.

La conducibilità termica è una misura della facilità con cui l'energia termica fluisce per diffusione; essa può esser misurata, per esempio nel caso dell'aria, con un semplice esperimento. Consideriamo un volume d'aria posto tra due lamine; la lamina superiore si trova alla temperatura di 330 °K e quella inferiore a 300 °K *(si veda la figura in alto a pag. 96)*. Le molecole urtando sulla lamina superiore acquistano un'energia cinetica media che è proporzionale alla temperatura di tale lamina; lo stesso accade per le molecole che urtano sulla lamina inferiore. In altre parole, le molecole che si trovano nella parte superiore del volume d'aria hanno in media un'energia cinetica maggiore (10% in piú) di quelle che si trovano nella parte inferiore. In uno stato stazionario il flusso di molecole verso l'alto è esattamente uguale al flusso di molecole verso il basso (poiché il numero totale di molecole tra le due lamine è fisso); questa condizione equivale a dire che la quantità di moto totale delle molecole che si dirigono verso l'alto è uguale alla quantità di moto totale delle molecole che si dirigono verso il basso.

Sebbene arrivino alla lamina superiore tante molecole quante sono quelle che la lasciano, l'energia cinetica delle molecole che si muovono verso il basso è maggiore di quelle che si muovono verso l'alto. In definitiva esiste quindi un flusso non nullo di energia

Un grafico tipico della temperatura per l'esperimento descritto nella figura in basso della pagina a fronte è disegnato qui per successivi intervalli di tempo dopo che il calore è stato applicato; i numeri vicini alle curve indicano il tempo misurato in secondi. Nell'istante di applicazione del calore l'estremità mostra un considerevole aumento della temperatura; altrove la temperatura è ancora a 0 °C. Un secondo dopo l'energia termica si diffonde attraverso il bastone e il grafico della temperatura comincia ad appiattirsi. Questo processo continua sino a che il bastone non ritorna alla temperatura uniforme di 0 °C.

Questo grafico dà l'andamento della temperatura se l'energia termica si propaga attraverso il bastone come un'onda. Il calore è contenuto entro una piccola regione dello spazio e viaggia con velocità costante. Impulsi di temperatura di questo tipo sono stati osservati in campioni di elio solido da ricercatori della Duke University e del National Laboratory di Los Alamos.

In un gas l'energia termica è condotta per diffusione dalle molecole che si muovono casualmente. Nel semplice schema, riprodotto per spiegare la conducibilità termica dell'aria, il calore si diffonde dalla lamina piú calda (*in alto*) verso la lamina piú fredda (*in basso*), poiché la maggior parte delle molecole che colpiscono la lamina piú calda la lasciano avendo aumentato la propria energia cinetica. Alla lamina piú fredda si ha il processo inverso e una certa quantità di energia è ceduta dalle molecole.

cinetica verso il basso, in breve, un flusso di calore. La spiegazione di questo fenomeno è una diretta conseguenza della teoria cinetica dei gas; la conducibilità termica fu infatti calcolata correttamente sin dai primordi di questa teoria in funzione del libero cammino medio, della velocità media delle molecole e del calore specifico del gas. La conducibilità termica in un gas è quindi un fenomeno completamente chiarito. Che si può invece dire sulla conduzione del calore in un solido?

1

2

3

4

In un solido cristallino ogni aumento di energia termica si manifesta di solito sotto forma di aumento di energia vibrazionale degli atomi attorno alle loro posizioni di equilibrio nel reticolo. Questo aumento è vincolato dalle forze interatomiche ad assumere la forma di onde che si propagano attraverso il campione. Questa situazione è analoga a quella che si verifica per un insieme di palline egualmente spaziate tra loro e disposte lungo un filo teso. Quando una pallina è leggermente spostata dalla sua posizione di equilibrio e viene poi lasciata libera, dal punto in cui è stata applicata la perturbazione si propaga un'onda a causa dell'accoppiamento elastico tra le palline.

In questa analisi ci limiteremo ai solidi elettricamente isolanti. Nei metalli e nei semiconduttori anche gli elettroni possono infatti contribuire alla conduzione del calore, e ciò rende l'analisi del flusso di calore in questi materiali troppo difficile per essere considerata in questa sede. Comunque il secondo suono non è stato osservato in questi materiali.

Per spiegare il flusso di calore nei solidi si deve esaminare dettagliatamente in che modo l'energia termica si manifesta nei cristalli, per esempio nel diamante o nell'elio solido. La piú importante caratteristica di questi solidi è la loro regolarità. Quando un fisico teorico cerca di calcolare qualche proprietà di un solido cristallino, la sua prima approssimazione è che il solido sia costituito da un insieme completamente regolare di atomi legati tra loro da forze che agiscono come molle. Gli atomi non sono a riposo ma vibrano con piccoli spostamenti attorno alle loro posizioni di equilibrio nel reticolo cristallino.

Supponiamo che un cristallo sia allo zero assoluto. Cosa accade se aumentiamo l'energia termica del cristallo aggiungendo calore? Allo zero assoluto tutti gli atomi vibrano negli stati con la piú bassa energia possibile, cioè con la piú piccola ampiezza permessa. Supponiamo ora di mettere una sorgente di calore in contatto con un'estremità del campione. Questa sorgente è naturalmente a una temperatura piú alta; in altre parole i suoi atomi vibrano piú violentemente. Non appena gli atomi della sorgente urtano quelli all'estremità del campione, le vibrazioni di questi ultimi aumentano. Le forze elastiche trasmetteranno allora l'eccitazione agli atomi piú interni del campione, cosicché tutti gli atomi vibreranno negli stati con energia piú alta. Quindi l'intero campione sarà a una temperatura piú alta.

L'aumento di energia termica si manifesta microscopicamente in un aumento dell'energia cinetica vibrazionale dei singoli atomi. Esso però non è casuale: vincolato dalla natura periodica ed elastica delle forze interatomiche è costretto ad assumere la forma di un'onda che viaggia attraverso il campione. Ciò può essere compreso meglio, per analogia, considerando un insieme di palline poste a intervalli regolari su un filo elastico (*si veda la figura qui accanto*). Se una delle masse è spostata di poco e quindi lasciata libera, essa non solo vibrerà ma trasmetterà il suo movimento anche alle masse piú vicine, che vibreranno a loro volta. In questo modo la perturbazione si propaga

I fononi, particelle associate alle onde termiche dalla teoria quantistica, possono esistere in un solido solo con determinate frequenze discrete. L'energia di ciascun fonone è data dall'equazione $E = h\nu$, dove E è l'energia, h la costante di Plank e ν la frequenza del fonone. L'energia termica totale in un cristallo è data dalla somma dell'energia di ciascuna frequenza per il numero di fononi con quella frequenza. Perciò in questo esempio l'energia totale è $6h\nu_1 + 6h\nu_2 + 2h\nu_3 + 3h\nu_4 + h\nu_5$.

lungo il filo e una fotografia fatta in un qualsiasi istante mostrerà che le masse assumono la forma di un'onda che si propaga dal punto in cui è stata applicata la perturbazione.

I solidi reali si comportano all'incirca nello stesso modo. L'aumento della ampiezza di vibrazione a una estremità è trasmesso agli atomi piú vicini e quindi a tutti gli atomi del cristallo nella forma di un'onda viaggiante. Quindi l'energia termica si manifesta come un'onda viaggiante nel solido.

È noto che la fisica classica non descrive correttamente fenomeni su scala atomica o molecolare, cosicché occorre dare una descrizione quantomeccanica di queste « onde termiche ». Un importante risultato della teoria quantistica è che le onde hanno anche proprietà corpuscolari e possono essere descritte come particelle; le particelle che rappresentano le onde termiche sono chiamate fononi, in analogia con i fotoni della luce e delle altre onde elettromagnetiche. Quindi un solido può essere descritto come una scatola contenente un gas di fononi.

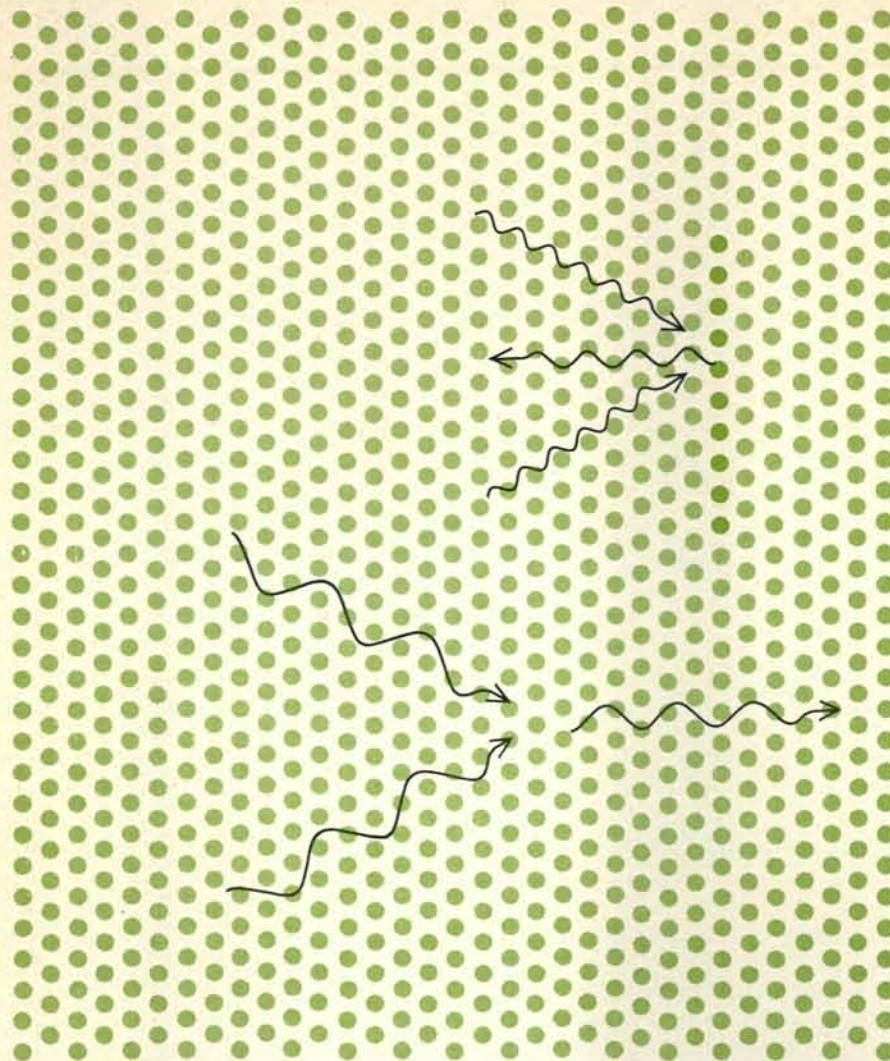
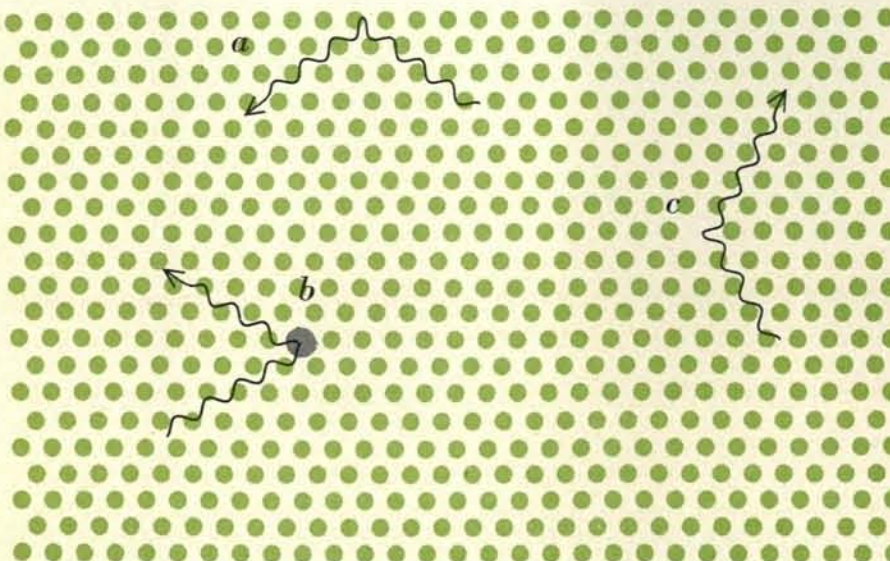
In meccanica quantistica alcune grandezze fisiche non possono variare su un intervallo continuo di valori, ma possono avere soltanto certi valori distinti; l'energia di un'onda termica, una delle sue proprietà piú importanti è quantizzata in questo modo. La differenza tra due qualsiasi valori permessi dell'energia dipende dalla frequenza dell'onda attraverso la condizione $\Delta E = \pm h\nu$, dove ΔE è la differenza d'energia, h è la costante di Plank e ν è la frequenza dell'onda. Il segno + corrisponde a un aumento in energia e quindi, nel linguaggio quantistico, alla creazione di un fonone, e similmente $\Delta E = -h\nu$ significa che un fonone di energia $h\nu$, viaggiante nella direzione dell'onda termica, è stato distrutto; quindi se si aggiunge energia termica a un solido ponendolo in contatto con una sorgente di calore, in realtà si eccitano onde termiche e quindi si creano fononi. Queste considerazioni rendono evidente una importante differenza tra le molecole di un gas reale e i fononi: essi possono essere creati o distrutti e quindi il loro numero non rimane necessariamente costante. Infatti il numero di fononi presenti in un solido è direttamente proporzionale alla sua temperatura.

Quando un fonone è distrutto, cosa accade della sua energia? Possono essere creati nuovi fononi in modo tale che l'energia sia conservata nell'intero processo? Esiste un tipo di interazioni fononiche che distruggono un fonone per crearne altri due e ne distruggono due per crearne uno; queste interazioni hanno luogo in presenza dei piani reticolari nel cristallo (*si veda la figura in alto nella pagina seguente*).

Una situazione analoga si ha nella riflessione della luce da parte di uno specchio; in questo caso fotoni incidono sullo specchio e vengono riflessi. Nei solidi due fononi incidono sui piani cristallini e formano un altro fonone la cui energia è pari alla somma delle energie dei fononi incidenti; la direzione del fonone risultante è deviata, e nel caso piú estremo il flusso di energia può invertire completamente la propria direzione.

Esistono quindi due tipi di interazione a tre fononi: uno chiamato processo « umklapp », o U, (dalla parola tedesca che significa rovesciamento) in cui la direzione del flusso di energia è alterata, l'altro in cui non interviene alcuna riflessione da parte dei piani cristallini. In questo processo normale, o N, il flusso di energia ha la stessa direzione prima e dopo l'interazione.

Perché un processo U sia possibile, almeno uno dei due fononi incidenti deve avere una lunghezza d'onda confrontabile con la distanza tra i piani cristallini; un tale fonone ha un'energia $h\nu$ vicina alla massima energia che può avere in un solido. Il numero di processi U aumenta all'aumentare della temperatura, poiché aumenta il numero di fononi presenti con grande energia (e piccola lunghezza d'onda). Anche il numero di processi N aumenta con la temperatura, poiché sono presenti fononi (di tutte le lunghezze d'onda) in maggior numero. In analogia con le molecole di un gas, si definisce libero cammino medio la distanza media percorsa da un fonone tra due urti successivi; nella figura a pagina 100 è mostrata la dipendenza del libero cammino medio dei fononi della temperatura per l'elio solido, che è però tipica dei solidi in generale.

Come in tutti gli altri fenomeni fisici, anche nei processi a tre fononi l'energia deve essere conservata: l'energia totale dei fononi uscenti deve esse-

L'interazione tra i fononi ha luogo sui piani reticolari. Nei processi normali, o *N* (*in basso*), due fononi si uniscono per formare un terzo fonone che si muove nella stessa direzione (oppure un fonone può essere distrutto per formare due nuovi fononi che si muovono nella stessa direzione). Nei processi « umklapp », o *U* (*in alto*), la direzione del flusso di energia è invertita dalla riflessione sul piano di atomi (*in colore pieno*).

I fononi sono diffusi da qualsiasi deviazione dalla periodicità perfetta della struttura cristallina. Per esempio, come mostra questa figura, la superficie del cristallo (*a*), difetti puntiformi su scala atomica (*b*) e vacanze del reticolo (*c*), diffondono i fononi.

re uguale all'energia totale dei fononi incidenti. Inoltre il fronte dell'onda termica emergente è ottenuto sommando gli spostamenti delle onde incidenti; quindi la direzione del fonone risultante non è arbitraria ma è determinata da questa condizione di interferenza. La direzione delle quantità fisiche viene descritta utilizzando i vettori; si può quindi associare alla direzione dei fononi un vettore la cui intensità è inversamente proporzionale alla lunghezza d'onda del fonone stesso; tale vettore, chiamato quasi-impulso, soddisfa, in un processo normale a tre fononi, le stesse leggi di conservazione dell'impulso reale tra le molecole in un gas.

Sinora abbiamo considerato i fononi come la rappresentazione corpuscolare di un'onda che attraversa un cristallo perfettamente periodico; un cristallo di questo tipo in realtà non esiste e ogni deviazione dalla periodicità deforma l'onda o, in altre parole, diffonde i fononi.

L'interruzione piú drastica della periodicità interviene sulla superficie di un solido e infatti i fononi sono grandemente diffusi alla superficie. Tutte le imperfezioni del cristallo, per esempio vacanze e dislocazioni, impurezze chimiche (un atomo di una specie chimica diversa da quella del reticolo ospite) e infine impurezze isotopiche (un atomo della stessa specie chimica ma di massa diversa da quella degli atomi del reticolo ospite) causano diffusione dei fononi (*si veda la figura in basso in questa pagina*).

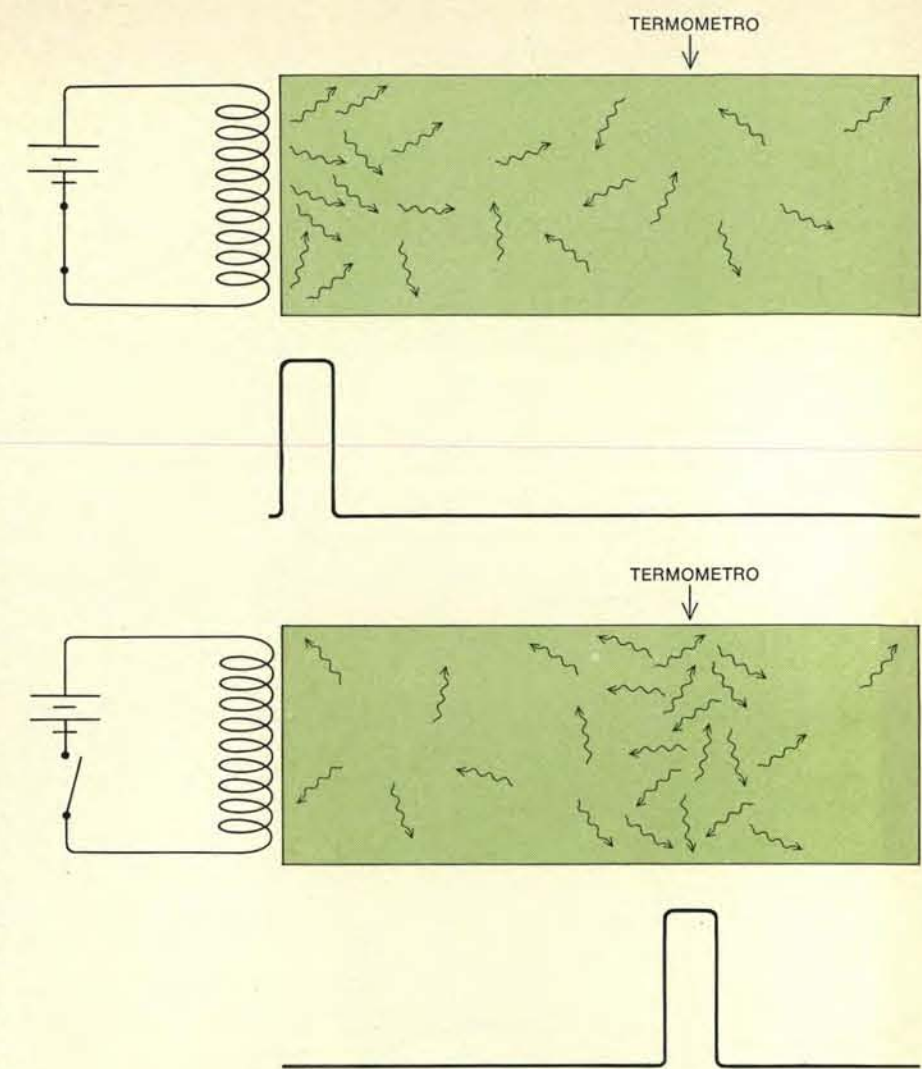
I processi di diffusione e i processi umklapp alterano la direzione del flusso di energia e sono chiamati processi resistivi poiché offrono resistenza alla conduzione termica; la conducibilità termica di un solido dipende dal libero cammino medio per i vari processi resistivi. I processi normali non causano resistenza termica, ma distribuiscono solamente l'energia tra i fononi.

La discussione sin qui fatta ci porta a concludere che la descrizione quantistica di un solido cristallino è simile alla descrizione classica di un gas: in un gas le molecole si muovono con differenti energia cinetica e impulso, in un solido i fononi si muovono con differenti energie e quasi-impulso.

Tutti sanno che il suono è un fenomeno ondulatorio e che può propagarsi attraverso un gas come l'aria; per ragioni storiche questa onda è chiamata primo suono. Se un altoparlante riproduce la musica di un disco, il cono dell'altoparlante stesso vibra con la frequenza necessaria a riprodurre nell'aria la musica che arriva alle nostre orecchie sotto forma di onde acustiche. Le frequenze della musica variano in un intervallo che va da 20 a 20 000 hertz (cicli al secondo) e le onde acustiche attraversano l'aria a una velocità di 350 metri al secondo, cosicché una lunghezza d'onda tipica (corrispondente a una frequenza di 1000 hertz) misura 35 centimetri. Il punto fondamentale, per il comportamento ondulatorio del suono, è costituito dal fatto che questa lunghezza d'onda è molto maggiore del libero cammino medio di una molecola nell'aria; ciò significa che la variazione di pressione dell'aria (e un'onda acustica è appunto un movimento attraverso l'aria della variazione di pressione) non si propaga per mezzo delle singole molecole ma per effetto di molte molecole; in altre parole le molecole si urtano cosí frequentemente tra loro che per distanze dell'ordine del centimetro si può trascurare la granulosità microscopica del gas: l'onda acustica si propaga come in un mezzo elastico continuo.

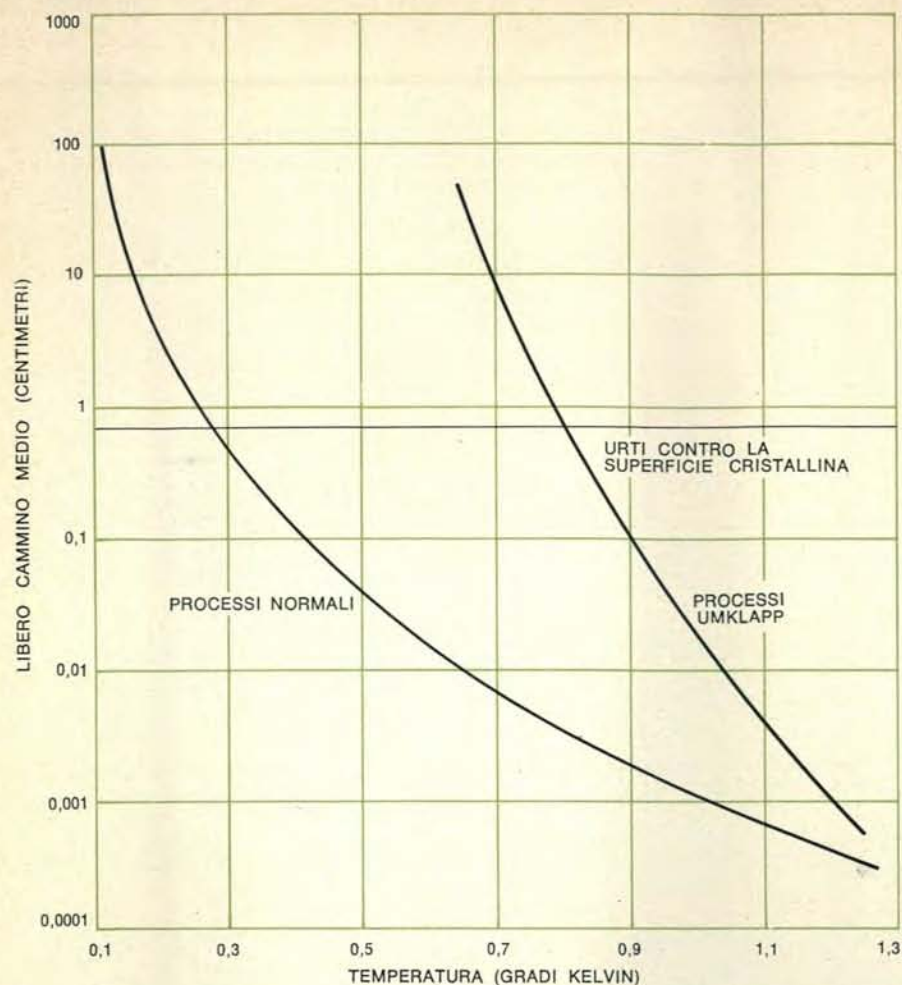
Generalmente il suono, e naturalmente la musica, non è un'onda di durata illimitata; per esempio se un violinista pizzica una corda, un impulso di energia acustica viaggia attraverso l'aria. Quando il cono dell'altoparlante preme sulle molecole che lo colpiscono, l'impulso medio delle molecole stesse (e quindi la loro energia cinetica media) aumenta; queste a loro volta urtano, a una distanza dall'altoparlante pari a molti liberi cammini medi, con altre molecole, trasmettendo loro l'aumento di impulso e di energia cinetica. Anche la densità delle molecole (il loro numero al centimetro cubo) aumenta nell'immediata vicinanza del cono dell'altoparlante, e ciò, tenuto conto dell'aumento di energia cinetica, comporta un aumento della pressione. Inoltre l'aumento dell'energia cinetica molecolare media corrisponde a una temperatura del gas leggermente piú alta; in altre parole, il suono ordinario, o primo suono, non è solamente un'onda che trasporta una variazione di pressione ma può dar luogo anche a piccole variazioni della temperatura.

In conclusione la propagazione del primo suono presenta le seguenti caratteristiche: 1) il primo suono può propagarsi in un gas, quando le molecole urtano cosí frequentemente l'una contro l'altra che la lunghezza d'onda del primo suono è molto piú grande del libero cammino medio; 2) le interazioni tra le molecole nel gas obbediscono alle leggi meccaniche della fisica, cioè alle leggi di conservazione dell'energia e dell'impulso; 3) il primo suono è un'onda e propaga simultaneamente variazioni di pressione e temperatura con perdita di energia molto piccola.

Questo processo è molto diverso dalla diffusione della temperatura, che non comporta variazioni di pressione.

In questa figura è mostrata la distribuzione dei fononi che si ha in un campione riscaldato in due istanti successivi durante la propagazione del secondo suono. Dapprima il riscaldatore viene acceso per una frazione di secondo (normalmente circa un milionesimo di secondo) e vengono cosí creati molti fononi (*in alto*). Poco tempo dopo che il riscaldatore è stato spento, l'impulso di fononi si è spostato verso destra (*in basso*).

Generalizzando le idee introdotte prima, si deduce che in un sistema di particelle interagenti l'una con l'altra in modo tale che energia e impulso (o quasi-impulso) siano conservate, le onde possono propagarsi se la loro lunghezza d'onda è molto maggiore del libero cammino medio.

Questa possibilità è stata suggerita indipendentemente da Laszlo Tisza nel 1938 e da Lev D. Landau nel 1941 nel caso dell'elio liquido superfluido a temperature inferiori a 2,2 °K. Landau suggerí che il comportamento dell'elio superfluido è descritto esattamente da un gas di particelle che egli chiamò « eccitazioni elementari » del liquido. Ci sono due tipi di eccitazioni nell'elio liquido superfluido: fononi e rotoni. Queste particelle possono essere create e distrutte, e il numero presente di ciascun tipo di esse dipende solamente dalla temperatura. Infatti, quando sono in equilibrio tra loro, dalla densità delle eccitazioni elementari si ricava immediatamente la temperatura dell'elio. I fononi e i rotoni interagiscono tra loro e hanno un libero cammino medio che dipende a sua volta dalla temperatura ed è molto minore di un millimetro a temperature superiori a un grado Kelvin.

Questo insieme di eccitazioni, o particelle, soddisfa alle condizioni necessarie per la propagazione delle onde, cioè gli urti tra le eccitazioni sono molto frequenti sopra un grado Kelvin (come gli urti tra le molecole nell'aria con un piccolo libero cammino medio),

Le condizioni per il secondo suono nell'elio solido con una densità di 0,205 grammi al centimetro cubo possono essere ottenute da questo grafico che rappresenta il libero cammino medio dei fononi in funzione della temperatura per i vari tipi di processi che li diffondono. A questa densità l'elio solido ha una struttura cristallina esagonale compatta; le curve mostrate nel grafico si riferiscono a un cristallo di diametro di circa 5 mm. Gli urti alla superficie cristallina e i processi umklapp sono processi resistivi; la diffusione normale è un processo non resistivo. Le condizioni per il secondo suono (il libero cammino medio deve essere molto maggiore del libero cammino medio non resistivo) sono soddisfatte per questo campione tra 0,4 e 0,9 °K.

e in questi urti energia e impulso sono conservati (come negli urti molecolari nell'aria). Landau predisse l'esistenza di un'onda nelle eccitazioni elementari (proprio come il primo suono è un'onda nelle molecole dell'aria) e la chiamò secondo suono; questa è un'onda di temperatura poiché a una variazione nella densità delle eccitazioni molecolari corrisponde una variazione nella temperatura.

I solidi differiscono dall'elio superfluido in quanto le eccitazioni elementari – fononi – sono soggette a interazioni resistive che alterano la direzione del flusso di calore, cioè non conservano il quasi-impulso. Nella maggior parte dei solidi esistono infatti difetti puntiformi, come le impurità chimiche o isotopiche, oppure imperfezioni cristalline, come le dislocazioni, che diffondono i fononi, cosicché le interazioni resistive hanno luogo piú frequentemente dei processi normali.

Consideriamo ora il flusso di calore pensando al solido come a una scatola contenente un gas di fononi. Nel punto in cui si applica un impulso di calore, si creano fononi che sono deviati dalla direzione di flusso a causa delle interazioni resistive. Il gas di fononi trasporta quindi l'energia termica per diffusione.

Supponiamo di avere un campione molto puro con una perfetta struttura cristallina, e supponiamo che i processi *U* e gli altri processi resistivi intervengano molto raramente mentre i processi *N* intervengano molto frequentemente; in altre parole, il libero cammino medio, confrontato con lunghezza del solido, sia piccolo per gli urti che conservano l'impulso, mentre sia grande per le interazioni resistive. Questa condizione è soddisfatta per l'elio solido per temperature tra 0,4 e 0,9 °K (*si veda la figura in questa pagina*). Il gas di fononi è in questo caso molto simile all'aria eccitata con un cono d'altoparlante. Un impulso di calore applicato a una estremità di un campione crea fononi, e questa fluttuazione nella loro densità si propaga come un'onda. Poiché il numero di fononi dipende strettamente dalla temperatura, il secondo suono si propaga nel gas di fononi.

Perché, allora, si osserva in pratica quasi solamente la diffusione termica? È molto difficile ottenere sostanze in forma sufficientemente pura in modo tale che sia soddisfatta la condizione sul libero cammino medio per i processi resistivi; le impurezze diffondono in maniera tale i fononi che la loro presenza anche in piccole quantità causa la diffusione termica. Viceversa l'elio solido può essere ottenuto in una forma estremamente pura. Tutte le impurezze chimiche sono congelate alla temperatura dell'elio liquido, cosicché il solido ricavato da questo liquido è molto puro; dei due isotopi stabili dell'elio, l'elio-3 è estremamente raro in natura, cosicché le impurezze isotopiche non sono importanti nei campioni naturali di elio-4. D'altra parte, l'elio-3 può essere ricavato artificialmente come prodotto del decadimento radioattivo del tritio, e poiché la differenza di massa relativa tra elio-3 e elio-4 è molto grande, le impurezze di elio-4 possono essere facilmente eliminate. La tecnica per produrre cristalli singoli di elio privi di difetti è pertanto molto sviluppata.

Finora l'elio rimane l'unico solido in cui sia stato osservato il secondo suono. Possiamo sperare di ottenere il secondo suono in altri solidi? La difficoltà maggiore è quella di ottenere sostanze sufficientemente pure sia chimicamente sia isotopicamente. John Rogers e Robert O. Pohl della Cornell University hanno osservato un'indicazione del secondo suono in un cristallo di fluoruro di sodio accuratamente preparato, ma la misurazione è estremamente difficile. Tuttavia con il miglioramento delle tecniche di purificazione dei materiali, dobbiamo aspettarci di osservare le onde termiche in altre sostanze, essendo raggiungibili in molti laboratori le basse temperature necessarie.

Il secondo suono è fondamentale tanto quanto la diffusione; la sua assenza generale è dovuta al fatto che la maggior parte dei solidi sono fondamentalmente impuri e che gli esperimenti non sono eseguiti a una temperatura sufficientemente bassa.

GIOCHI MATEMATICI

di Martin Gardner

Come risolvere dei quiz tracciando la traiettoria di una palla che rimbalza

Nel corso della storia la palla elastica è stata un elemento indispensabile per una straordinaria varietà di sport da praticare sia all'aperto sia al chiuso. I giochi che ne fanno uso vanno da quello semplice del bambino che fa rimbalzare una palla di gomma (uno, due, tre,...), a sport come il tennis, la palla a mano e il biliardo in cui è essenziale per un buon giocatore la capacità di valutare angoli di incidenza e di riflessione.

È risaputo che matematici e fisici hanno una gran passione per il biliardo. È facile capire il perché: i rimbalzi ben ponderati su un tavolo perfettamente liscio possono essere calcolati con precisione. Lewis Carroll, che insegnò matematica all'Università di Oxford, si divertiva a giocare a biliardo e soprattutto su un tavolo *circolare* che egli stesso si era costruito. È un pezzo da collezionista molto raffinato la prima edizione di un articoletto di due pagine, pubblicato da Carroll nel 1890 e mai ristampato, che spiega le regole di questo gioco.

Centinaia di problemi-passatempo riguardano rimbalzi di palle elastiche all'interno di perimetri di varie forme. Considerate, per esempio, il seguente vecchio quiz: avete due contenitori di capacità di 7 e 11 litri rispettivamente. Accanto avete una gran vasca piena d'acqua. Usando solo i due contenitori (ed escludendo ogni trucco tipo il graduare i contenitori o inclinarli in modo da ottenere parti frazionarie) come potete separare esattamente due litri d'acqua?

Si può dare una risposta a questo problema o per tentativi, o applicando vari procedimenti algebrici. Ma che cosa ha a che fare tutto ciò con le palle che rimbalzano? Sorprendentemente, i quiz di questo tipo sulla misura dei liquidi si possono risolvere con facilità disegnando i percorsi di palle che rimbalzano all'interno di tavoli *romboidali*! Le sponde di tali tavoli si rappresentano meglio su carta millimetrata isometrica, cioè su una carta con un reticolo di triangoli equilateri. In tal caso i lati della tavola si rappresentano come segmenti lunghi 7 e 11 unità (*si veda la figura in questa pagina*). Lungo l'asse orizzontale si legge la quantità di acqua nel contenitore da 11 litri mentre sull'asse verticale è riportata la quantità di acqua nel contenitore da 7 litri.

CONTENITORE DA 11 LITRI	11	4	4	0	11	8	8	1	1	0	11	5	5	0	11	9	9	2
CONTENITORE DA 7 LITRI	0	7	0	4	4	7	0	7	0	1	1	7	0	5	5	7	0	7

Grafico e soluzione in 18 mosse di un problema sui liquidi.

Per usare il grafico, si immagini una palla che sia nel punto 0, nell'angolo in basso a sinistra. La palla si muove verso destra lungo la base del romboide fino a che batte sulla sponda destra nel punto di ascissa 11, sulla linea di base: questo significa che il contenitore da 11 litri è stato riempito mentre quello da 7 è ancora vuoto. Dopo esser rimbalzata sulla sponda di destra la palla si muove verso l'alto a sinistra, finché batte contro la sponda in alto nel punto di ascissa 4 e di ordinata 7. Questo tratto del grafico sta a indicare che si sono travasati sette litri d'acqua dal contenitore da 11 in quello da 7, lasciandone solo 4 nel più capace.

Se si continuano a seguire i rimbalzi della palla fino a che questa batte in un punto contrassegnato col numero 2, registrando i vari movimenti si otterrà la risposta in 18 mosse mostrata sotto il grafico. Le frecce oblique indicano che l'acqua viene travasata da un contenitore all'altro; quelle verticali mostrano che il contenitore da 7 è stato vuotato nella vasca, o che quello da 11 è stato riempito.

Ma questa è la soluzione più breve? No; un altro procedimento è quello di iniziare riempiendo per primo il contenitore da 7 litri. Ciò si può rappresentare graficamente facendo partire la palla dal punto 0 e spingendola verso l'alto lungo il bordo sinistro. Se il lettore traccia il grafico fino a che la palla arriva a un punto contrassegnato con 2, troverà che la sua palla calcolatrice arriva a una soluzione in 14 mosse, il minimo.

Con un po' di abilità si potrebbero ideare dei calcolatori a palla elastica per qualunque quiz sui liquidi in cui non ci siano più di tre contenitori. Consideriamo il più antico tra tutti i problemi che riguardano tre contenitori, che risale a Nicola Fontana, il matematico italiano del sedicesimo secolo che fu soprannominato Tartaglia. Si riempie con acqua un contenitore da 8 litri. Per mezzo di due bottiglie vuote che contengono 5 e 3 litri rispettivamente, si devono suddividere gli 8 litri in due parti uguali che siano nei due contenitori più capaci. Il grafico che dà la soluzione di questo problema è mostrato in alto a sinistra. Il contenitore da 8 litri è rappresentato da un segmento parallelo alla diagonale del romboide. La palla parte come prima dall'angolo 0. È facile tracciare un percorso che dà la soluzione minima, che richiede sette operazioni.

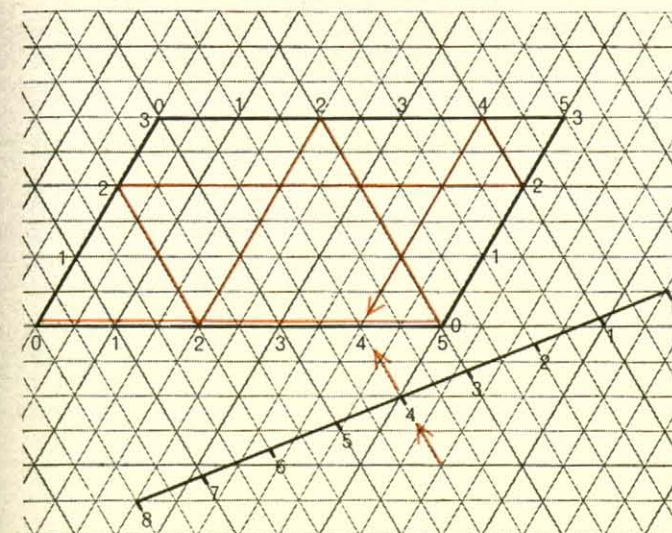
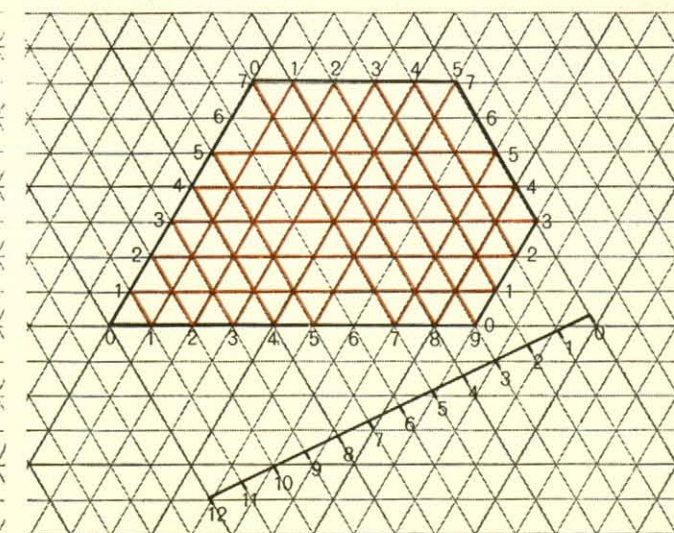



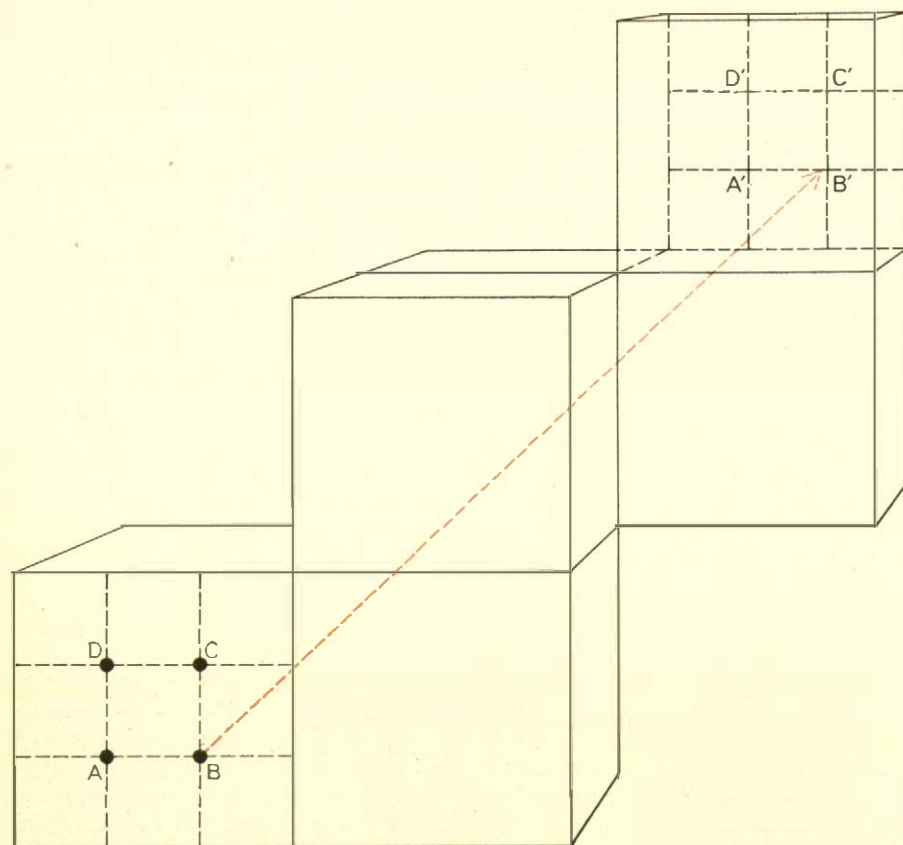
Grafico del problema di Tartaglia. *Grafico per contenitori da 7, 9 e 12 litri.*

Quando le capacità dei due contenitori più piccoli non hanno divisore comune e quella del terzo è maggiore o uguale alla somma della capacità degli altri due, è possibile dosare una qualunque quantità di liquido compresa tra un litro e la capacità del contenitore medio. Per esempio, con contenitori da 15, 16 e 31 litri si potrebbe separare una qualunque quantità tra uno e 16 litri. Questo però non è possibile se i due contenitori più piccoli hanno un divisore comune. In un grafico per un problema con contenitori da 4, 6 e 10 litri, la palla non rimbalzerà mai in punti di coordinate dispari, mentre con contenitori da 3, 9 e 12 litri si potranni separare solo 3, 6 o 9 litri. (In entrambi i casi sarà possibile separare solo multipli del divisore comune). Se poi la capacità del contenitore più grosso è *minore* della somma della capacità degli altri due, allora ci sono ulteriori limitazioni. Per esempio con contenitori da 7, 9 e 12 litri è necessario eliminare un angolo dal grafico romboidale (*si veda la figura in alto a destra*). La palla rimbalzando potrà misurare una quantità qualunque di liquido tra 1 e 9, tranne 6. Sebbene 7 e 9 non abbiano divisori comuni la scarsa capacità del terzo contenitore non dà la possibilità di ottenere 6.

Questo metodo si può applicare anche quando la capacità del contenitore più grosso è maggiore della somma della capacità degli altri due. Il lettore potrà divertirsi applicandolo alla seguente variazione del problema di Tartaglia.

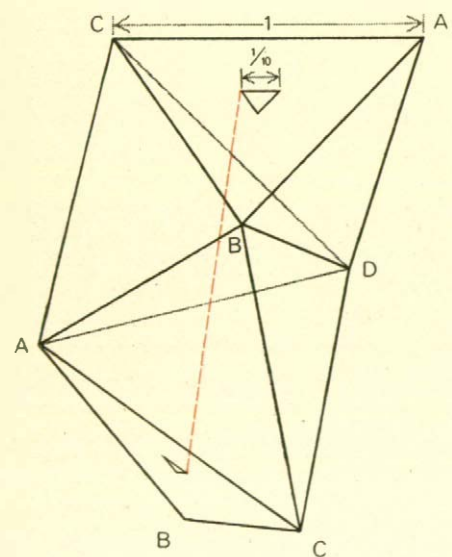
Alcuni soldati riuscirono a « impadronirsi » di un barilotto da 10 litri di birra. Essi naturalmente ne assaggiarono un po', servendosi di recipienti da 3 e da 5 litri. Il resto della birra fu riportato al campo suddiviso in tre parti uguali, una nel barilotto e le altre due nei recipienti. Quanta birra bevettero e come fecero a suddividere quella che era rimasta in tre parti uguali? La soluzione migliore è quella che richiede un minor numero di mosse. Ogni mossa che riguarda l'operazione del bere, si riferisce a un numero intero di litri, e si assume anche che non ci sia birra versata fuori dai recipienti. Il prossimo mese daremo la risposta e mostreremo come la si possa ottenere col metodo delle palle elastiche.

Il lettore troverà divertente provare con contenitori di varie capacità, usando il metodo delle palle per scoprire tutto quello che si può fare con esse.

Un tipo diverso di problemi risolubili col metodo delle palle che rimbalzano è quello di trovare dei cammini chiusi (cicli) lungo i quali la palla possa continuare a rimbalzare all'interno di un poligono, sempre facendo lo stesso percorso e toccando ogni lato solo una volta in ogni ciclo. Questi problemi si possono risolvere servendosi della tecnica della riflessione speculare. Un tavolo a forma di triangolo equilatero può fornirci un semplice esempio. Si supponga di porre la palla nel punto mostrato nella parte superiore della figura qui accanto. Si vuole spingerla verso il lato *BC* in modo che vi rimbalzi, colpisca il lato *AC* e ritorni nella sua posizione iniziale sul lato *AB*. Il cammino che dovrà percorrere si può trovare facilmente per mezzo di uno specchio che rifletta il triangolo due volte, ogni volta lungo un lato diverso, come è indicato in figura dalle linee tratteggiate. La linea retta che va dalla palla al punto dell'ultimo specchio in cui essa si riflette, ci fornisce la traiettoria richiesta. Si devono solo riportare i due triangoli tratteggiati sull'originale. Questo si può fare materialmente disegnando i triangoli su carta da lucido e tracciando una linea sottile che vada dalla palla alle sue immagini. Si ritagli la figura e si pieghino i triangoli in modo da ottenerne uno solo. Ciascun triangolo ripiegato è un'immagine speculare dell'originale. Tenendo in controluce i triangoli sovrapposti si vedrà il cammino ciclico come nella parte inferiore

Cammino di riflessione in un triangolo equilatero.

Soluzione del problema di Lewis Carroll nel cubo.

dell'illustrazione. Se aggiungiamo l'ulteriore limitazione che tutti i segmenti di tale percorso debbano essere della stessa lunghezza, è facile vedere che può esservi un'unica soluzione: il cammino dato dai segmenti che congiungono i punti medi dei tre lati.

Con lo stesso procedimento si possono trovare cammini ciclici all'interno di un qualunque altro poligono, per esempio di un quadrato.

Soluzione al problema nel tetraedro.

A questo punto sorgono due questioni interessanti. È possibile trovare cammini ciclici con tratti tutti uguali tra loro all'interno del cubo e del tetraedro che sono i solidi corrispondenti al quadrato e al triangolo equilatero? Si considera la palla come se fosse una particella elastica ideale (o un raggio di luce in un solido con superfici interne speculari) che percorra, in assenza di gravità, dei tratti rettilinei e che rimbalzi contro i lati nella solita maniera secondo angoli di incidenza e di riflessione uguali su un piano perpendicolare alla faccia su cui rimbalza. La palla deve colpire ciascuna faccia una sola volta durante l'intero ciclo e deve percorrere la stessa distanza tra due rimbalzi consecutivi. (Non si può considerare la palla che colpisca un lato o un angolo come se avesse colpito le facce che hanno quell'angolo o quel lato in comune; altrimenti il problema del cubo si potrebbe risolvere con una palla che si muova avanti e indietro tra due angoli diametralmente opposti.)

Warren Weaver, in uno dei suoi numerosi articoli su Lewis Carroll, ha rivelato che il problema del cubo è stato trovato tra le note matematiche inedite di Carroll. Questo tipo di problema è proprio tra quelli che dovevano interessare l'ideatore del bigliardo rotondo. In realtà l'idea di giocare a bigliardo all'interno di un « tavolo » cubico non è cosí assurdo come si potrebbe pensare. Con le gigantesche stazioni spaziali che ci saranno probabilmente ben presto, non ci vuole una grande mente profetica per prevedere lo sviluppo di una gran varietà di sport tridimensionali, favoriti dalla mancanza di gravità. La partita a bigliardo si adatta perfettamente a essere giocata in una stanza rettangolare con pareti, pavimento e soffitto elastici e con buche negli angoli. Naturalmente ci sarebbero delle difficoltà. La resistenza dell'aria offre un attrito molto minore di quello della superficie di feltro di un tavolo da bigliardo; inoltre sarebbe difficile ripararsi dalle palle che volerebbero qua e là in direzioni casuali come le molecole di un gas in equilibrio termico!

Ma torniamo al problema di Carroll. La tecnica della riflessione usata con i quadrati, si può applicare anche ai cubi; con cinque immagini riflesse si può ottenere il percorso voluto che è rappresentato dalla linea in colore nell'illustrazione in alto in questa pagina; si ha cosí uno dei quattro percorsi diversi, identici nella forma, che risolvono il problema (se tutte e sei le facce del cubo fossero suddivise in nove quadrati piú piccoli, ciascun percorso toccherebbe le facce in uno dei vertici del quadrato centrale). Se si immagina il cubo come un solido formato da 27 cubetti piú piccoli, ci si accorge che ogni tratto del percorso è una diagonale di un cubetto. Ogni tratto ha perciò una lunghezza di $1/\sqrt{3}$ su un cubo unitario. La lunghezza totale del percorso è perciò $2/\sqrt{3}$.

Per quanto io sappia, Roger Hayward, che collabora alla rubrica *Scienza in casa* di questa rivista, è stato il primo a trovare questa soluzione. La forma del percorso è nota a coloro che si occupano di chimica organica come un « esagono a sedia » che si riscontra spesso nei composti del carbonio, come il cicloesano, in cui sei atomi di carbonio sono collegati ad anello, con altri atomi attaccati all'esterno dell'anello. « È interessante notare – scrive B. M. Oliver – che il cammino appare come un rettangolo di dimensioni 1×2 in ciascuna delle proiezioni del cubo, prese perpendicolarmente a una faccia, come un rombo nelle tre proiezioni isometriche prese parallelamente alla diagonale del cubo, come un esagono regolare nella quarta prospettiva isometrica. Questa è una strana figura, ma rappresenta proprio il percorso delle palle elastiche! »

Un cammino ciclico simile a questo

Strategia per il Black cinque per cinque.

è stato trovato per un tetraedro nel 1962, ancora da Hayward. È facile far riflettere un tetraedro tre volte (*si veda la figura in basso a fronte*) e ottenere un percorso ciclico che tocchi ogni faccia una sola volta; la difficoltà consiste nel trovare un cammino ciclico con lati tutti uguali. Uno è quello mostrato dalla linea in colore. Ci son altri tre simili percorsi, che toccano tutte le facce del solido in un angolo del piccolo triangolo equilatero che sta al centro di ogni faccia. Il lato di questo triangolo è 1/10 di quello del tetraedro di lato unitario. Ogni tratto del percorso ha quindi una lunghezza di $\sqrt{10}/10$ o 0,31622777, che dà 1,2649 come lunghezza totale del percorso.

Hayward costruí un bel modello di acetato in cui un filo di nylon rappresenta il percorso della palla elastica (o del raggio luminoso), dopo che i quattro tetraedri sono stati reinseriti uno nell'altro. Egli ritagliò le facce da lastre di acetato e dopo aver fatto quattro piccoli fori nei punti appropriati, le incollò lungo i bordi. Prima di incollare anche l'ultima faccia fece passare il filo attraverso i buchi dalle altre tre e lo fissò con un nastro adesivo dal di fuori; i due estremi liberi del filo furono fatti passare attraverso il buco della quarta faccia, che a questo punto poté essere incollata alle altre. Dopo aver teso il filo tirandolo dall'esterno, otturò i buchi con una goccia di acetone misto a mastice e tagliò via i nodi e i capi del filo rimasti fuori. Un modello simile si può fare anche per il cubo. In entrambi i casi si possono usare fili di diverso colore per rappresentare i diversi percorsi.

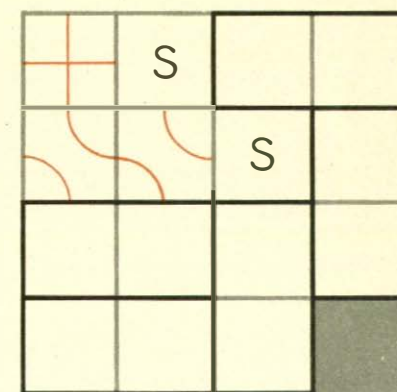
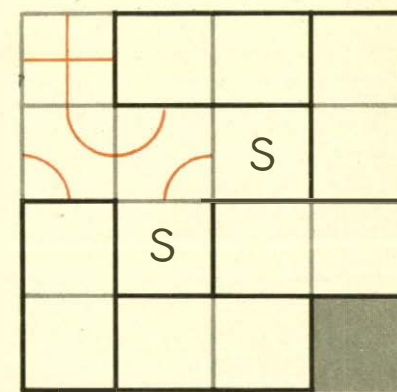
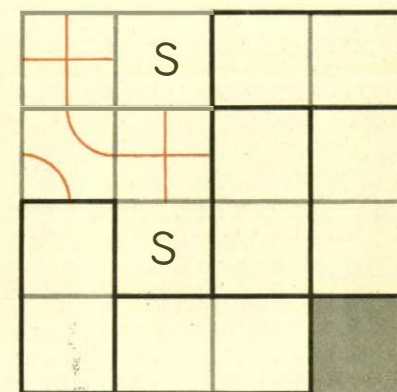
Ecco le risposte alle domande poste il mese scorso. Per quanto riguarda il gioco del soldato nel quale a Nero è dato il privilegio di iniziare il gioco da qualsiasi casella vuota, il problema è stato risolto per la prima volta dall'olandese Frederik Schuh nel 1943. Bianco riesce sempre a intrappolare Nero se gioca razionalmente. Non possiamo dare qui un'analisi completa del gioco, ma la tavola seguente mostra le risposte vincenti di Bianco alle sei differenti mosse di apertura di Nero (i numeri e le lettere sono quelli usati il mese scorso):

Nero	Bianco
2	A35
4 (o 6)	A15 (o A35)
5	123
7 (o 9)	A15 (o A35)
8	A15
B	123

Il gioco topologico di Black, del quale si doveva trovare la strategia vincente, è vinto dal primo giocatore se il numero totale di caselle è dispari mentre è vinto dal secondo se il numero totale è pari.

Quando il gioco si svolge su una scacchiera con numero dispari di caselle, per esempio una scacchiera cinque per cinque, la strategia del primo giocatore è quella di immaginare la scacchiera, eccettuata la casella nell'angolo in basso a destra, ricoperta da pedine di domino (*si veda la figura in alto in questa pagina*); la disposizione effettiva delle singole pedine è irrilevante. Ogni mossa del secondo giocatore inizia il percorso su una nuova pedina e il primo giocatore deve giocare in modo che il percorso *rimanga su tale pedina*. Questo costringe il secondo giocatore a completare la pedina e quindi a far iniziare il percorso su una nuova pedina: è ovvio che alla fine il secondo giocatore sarà costretto a raggiungere un bordo o la casella nell'angolo in basso a destra.

Su una scacchiera con numero pari di caselle la strategia che porta alla vittoria il secondo giocatore è molto piú complicata. La scacchiera deve essere ancora immaginata ricoperta da pedine del domino eccettuate però le due caselle negli angoli in alto a sinistra e in basso a destra.

Poiché le due caselle non coperte hanno lo stesso colore (supponendo che la scacchiera sia colorata come quella degli scacchi), è chiaramente impossibile coprire tutte le altre con pedine da domino: vi saranno sempre due celle non coperte dello stesso colore. Berlekamps, che ha analizzato il gioco, ha chiamato queste due caselle scoperte la « pedina spezzata ». La pedina spezzata viene individuata nel seguente modo: il secondo giocatore fa la sua prima mossa come è mostrato nel primo dei disegni qui sotto. Tale mossa costringe il primo giocatore a giocare nella seconda casella della diagonale principale uno qualsiasi dei tre simboli, come mostrano i tre disegni qui sotto. In ognuno dei casi la linea libera del simbolo usato collega due caselle dello stesso colore e sono proprio queste caselle, indicate con *S* nel disegno, che costituiscono la pedina spezzata. Le caselle restanti, esclusa al solito quella d'angolo, possono ora essere ricoperte con pedine da domino la cui disposizione è anche questa volta irrilevante. A questo punto il secondo giocatore vince con il metodo del domino spiegato prima.

Strategia per il Black quattro per quattro.